Neural Indicators of Inference and Recognition Processes in Language Comprehension

Dissertation zur Erlangung des Grades eines
Doktors der Kognitionswissenschaft (PhD in Cognitive Science)

eingereicht von
Uwe Friese, Dipl.-Psych.

an der Universität Osnabrück
Fachbereich Humanwissenschaften

Mai 2009

Acknowledgements

Throughout my work on this thesis I have received support from a number of people who should be mentioned here.

Naturally, this project could not have been completed without the backing of my supervisors Peter Bosch, Evelyn Ferstl, and Mark Greenlee. I list their names in alphabetical order because I am equally grateful for their knowledge, guidance, and words of encouragement. I am aware that the type of support they provided cannot be taken for granted. Furthermore, Carla Umbach encouraged me to join the Cognitive Science PhD-programme which proved to be an ideal platform for inter-disciplinary research and learning. I am also obliged to the presidential board of the University of Osnabrück for enabling me to finish the biggest part of this thesis during my time in Osnabrück.

My experiments were planned and carried out while I was research associate in the lab of Franz Schmalhofer who supervised this project in the beginning. I am thankful for his suggestions and support. Several students at the Institute of Cognitive Science assisted with, e.g., constructing text materials and with testing experimental procedures. Representative for them I want to thank Karin Pietruska, Carsten Griesel, Jan Peters, and Nadia Hristova.

MR-data for this thesis was acquired in the lab of Mark Greenlee at the University of Oldenburg, and I owe gratitude to Markus Raabe, Roland Rutschmann, Riklef Weerda, and Jale Özyurt. Roland and Markus were invaluable teammates, and I especially enjoyed our profound discussions in the operating room of the MR-scanner and in the "Banana-Bar". Cheers! Without them I would still battle against SPM. Markus deserves special mention for E-prime coding and his enthusiastic assistance throughout all stages of this project. I am extremely happy that finally our successful cooperation is also reflected in a journal publication of Experiment 1 (Friese, Rutschmann, Raabe, & Schmalhofer, in press).

The last years would have been significantly less enjoyable without serious and not so serious conversations with my colleague and friend Ho Ming Chow. Besides teaching me elementary Cantonese expressions, he was always prepared to discuss even the weirdest thoughts about language comprehension processes and functional neuroimaging experiments.

Finally, my deepest appreciation goes to my family who often had to cope with a stressed out and moody husband/father/son/brother who could not shake off the feeling that he had too little time.

To Monika, Yasmina, and Laura.

Summary

The human ability to make sense of language utterances is remarkable because of the complex interplay of a diversity of underlying cognitive processes. To date an increasing number of neuroimaging studies has tried to characterize the brain networks which accomplish language comprehension—sometimes with diverging results. In this research two functional magnetic resonance tomography experiments were conducted to identify core regions of language comprehension processes. The focus of the studies was on inferencing, i.e. the activation of information which has not been explicitly mentioned in a given utterance but which is somehow implied because of general world knowledge. The research strategy was two-fold. First, text materials were used which allowed to isolate inference processes from more basic language processes on the basis of the text representation model of van Dijk and Kintsch (1983). Second, two tasks—verification in Experiment 1 and recognition in Experiment 2—were assigned to the participants to selectively enhance or attenuate processing at different levels of representation.

In both experiments a network of brain areas was found to be active during language comprehension including areas all along the left superior temporal sulcus, the left lateral and medial prefrontal areas, as well as the right anterior temporal lobe and the posterior cingulate cortex. The results of Experiment 1 indicated that the dorsomedial prefrontal cortex was most prominently associated with inferencing in the context of the verification task. As expected, activity in this region was attenuated in Experiment 2 during recognition. No indications were found that the right hemisphere plays a particular role for inferencing as has been suggested by some authors.

The results of both experiments were discussed with respect to the neuroimaging literature on language comprehension and with respect to recent approaches to memory systems in the brain—particularly the episodic memory system. On the basis of this discussion a functional neuroanatomical model of inferencing was sketched.

Contents

Abbreviations of Brain Regions

AnG	angular gyrus
aTL	anterior temporal lobe
dIFG	dorsal inferior frontal gyrus
dlPFC	dorso-lateral prefrontal cortex
dmPFC	dorso-medial prefrontal cortex
IFG	inferior frontal gyrus
IPL	inferior parietal lobe
ITG	inferior temporal gyrus
LH	left hemisphere
MTG	middle temporal gyrus
mTL	mid temporal lobe
PCC	posterior cingulate cortex
pTL	posterior temporal lobe
RH	right hemisphere
SMA	supplementary motor area
SMG	supramarginal gyrus
STG	superior temporal gyrus
STS	superior temporal sulcus
TL	temporal lobe
TPJ	temporal parietal junction
vIFG	ventral inferior frontal gyrus
vlPFC	ventro-lateral prefrontal cortex
vmPFC	ventro-medial prefrontal cortex

A. Introduction

Research on language comprehension processes is confronted with a striking contrast: On the one hand, under most circumstances we understand a given utterance or a piece of written text without any particular effort. On the other hand, we need to acknowledge that many cognitive mechanisms are necessary to achieve comprehension. Even if we ignore the perceptual processes during the translation of the external auditory or visual stimuli to internally represented symbols, the list of important cognitive processes that play a potential role is still noticeable. Knowledge from semantic and episodic memory needs to be recruited. Executive working-memory related processes must guide attention and control memory retrieval. Moreover, reasoning processes, the motivations and goals of the comprehender, and, even more generally, the social, communicative setting in which language is produced and received have to be taken into account. All these complex processes take place in the human brain with such an amazing speed and efficacy that we are usually not aware of them.

In this thesis, the focus is on inference processes in language comprehension, or in more general terms, on the activation of specific parts of general knowledge by language utterances. To illustrate the type of inferences which are investigated here, consider the following short passage:

(1) *Coming home after running for two hours on a very hot day, Tom was exhausted.*
(2) *He reached into the fridge to get the bottle with orange juice.*
(3) *His heart was still beating heavily, and his hands were shaking.*
(4) *The bottle slipped out of his hand.*
(5) *It fell onto the kitchen floor.*
(6) *A piece of glas hit his* foot.

This text describes a chain of events which appear to be causally connected to each other. In (1) we learn that Tom is exhausted *because* a ran for two hours. As we know from experience that running for some time in hot weather is a plausible reason for being exhausted, the sentence is coherent. Moreover, the situation described in (1) allows us to infer further information about Tom‘s physiological and mental state. We can infer, e.g., that Tom is likely to be thirsty, which gives the reason for his actions in (2). We might also infer at this point that his heart is beating heavily, which is lateron mentioned in (3), or we could infer that he might sweat, which not stated in the text. After having read sentence (2), most readers will assume that Tom is thirsty and wants to drink something. It is important to note that this assumption is based on experience and general knowledge about similar situations—neither being thirsty nor drinking is mentioned explicitly in the text. The chain of events proceeds with the bottle slipping out of Tom ´s hand in (4). Again, our experience suggests a causal relationship between these events although other reasons might exist which could possibly cause Tom to drop the bottle. In the following, (4) allows to infer (5). Additionally, to understand sentence (6) readers need to infer that the bottle broke. Further inferences could be drawn after reading (6): The cut foot might be bleeding, orange juice might be spilled across the floor, Tom might be angry, and so on.

In the current studies, inferences are investigated which establish causal connections between events described by sentences. The basis for these inferences are the situations that are described in the sentence or discourse, and the comprehender´s knowledge about these or similar situations. The inferences provide information about antecedents or consequences of events which are not stated explicitly in the text. They are *causal* in the sense that they answer why certain events happen. In the current context, a broad conceptualization of causality is used which is not limited to physical effects. Instead, mental states which can explain behavior or which can be the result of a situation—e.g. being angry about a broken bottle—are also assumed to be causally related to events. In addition, the concept of inferences is much broader than in, e.g., formal logic.

While there is an extensive literature on behavioral studies of sentence and discourse level comprehension, cognitive neuroscience research in this field has only started. Functional neuroimaging methods uniquely provide a direct window to the brain processes which underly cognition and thereby ground and constrain cognitive theories (Wager, 2006). In addition to providing information about the location of cognitive functions, imaging data can inform about the polarity and the timecourse of brain responses (Ferstl, 2007a). This makes it possible to detect commonalities and differences between cognitive processes which cannot be revealed by behavioral measures. This thesis contributes to the growing number of neuroimaging studies investigating the neural basis of higher level language comprehension processes. In the following, the representations and processes involved in inference construction are specified. In two fMRI experiments these specifications are then tested for their validity in view of the functional neuroanatomy of comprehension processes.

The first part of this thesis introduces the most important theoretical concepts for the two experiments which are reported in later parts. This overview will be comparatively selective and tailored to the goal of providing the information that is necessary to understand the rationale of the current studies. Influential models of language comprehension are presented with respect to the underlying cognitive representations and processes. Additionally, some of the behavioral research methods and important results are summarized as they provide the background for the subsequent review of neuroimaging studies on language comprehension. In the second section two studies that used functional magnetic resonance tomography (fMRI) to map cognitive processes to brain structures are reported. While Experiment 1 utilized a verification task to identify the network of brain areas that supports inference processing, in Experiment 2 a recognition task was used to further subdivide the functional roles of the components of this network. In the last section of the thesis the implications of the current findings for functional-anatomical models of inference processing in language comprehension are discussed.

A.1. Language Comprehension and Inferences

Zwaan and Singer (2003) stated that “Almost every facet of comprehension is at least partly inferential” (page 100). The introductory text example illustrated some “inferences” that could be derived during reading. This chapter now provides a more systematic overview of theoretical concepts and models of language comprehension and inference processes that have stimulated a lot of research.

A.1.1. Levels of Representations and Coherence

A central aspect of language comprehension research is the question how language is mentally represented. Van Dijk and Kintsch (1983) assume that for understanding a text, readers construct a representation of the text itself as well as a representation of the situation the text is about. Their influential model distinguishes between representations of (a) the surface structure of the text, (b) the meaning of the text elements (propositions), and (c) the situation described by the text (situation model). These considerations were also crucial for the design of the current experiments as will be shown later in this chapter. In addition to these three levels it has been suggested that there are further representations at the communication level and the text genre level, emphasizing the fact that language comprehension usually takes place in communicative settings and serves certain purposes (Graesser, Millis, & Zwaan, 1997). Although these latter aspects can play an important role for language understanding, they are of minor importance for the current studies and are therefore not discussed any further.

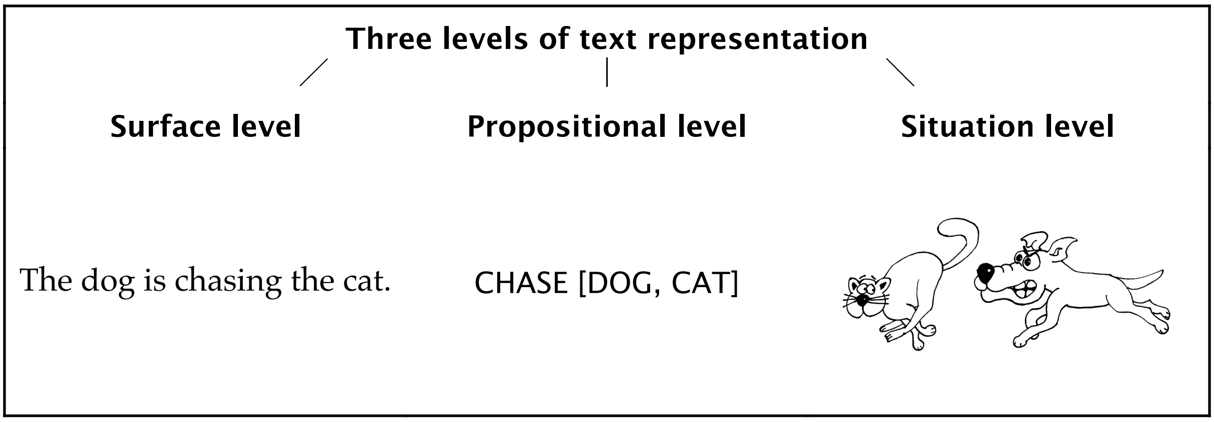


Figure 1: Illustration of three representational levels. At the surface level, the words that make up the sentence are represented. At the propositional level, meaning-based elements of the text are represented independent of the verbatim wording. The situation level representation contains information about the situation that is described by the text. (See text for further explanations.)

Figure 1 illustrates surface level, propositional level, and situational level representations of the sentence *The dog is chasing the cat*. Van Dijk and Kintsch (1983) assume that at the surface level the exact wording and syntactical structure that build the sentence are represented. This representation usually decays rapidly as the literal wording of a text is rarely important for comprehension.

The propositional level representation, often called the *textbase*, contains more abstract meaning representations of the words and facts explicitly mentioned in the text, while syntax and literal wording are not preserved. The use of the term *proposition* here differs from it's use in philosophy or linguistics. Propositions in the current context refer to units of meaningful content. CHASE[DOG, CAT] is meant to represent the information that the dog is chasing the cat. The predicate-argument notation is used to illustrate the independence of propositions from surface features of the text. For instance, it is assumed that the sentence "The dog is running after the cat" results in the same propositional representation as "The dog is chasing the cat" because the sentence meanings are identical. Thus, the textbase is constructed by decomposing a text into meaningful units, i.e. propositions, and it is assumed that it constitutes the semantic representation of the input text irrespective of the original wording. It has been shown that usually meaning-based representations of a text are stored for a longer time than surface representations of text (Bransford & Franks, 1971; Bransford, Barclay, & Franks, 1972).

Finally, the situation model represents the state of affairs the text is about. It is detached from the text structure, and based on the reader´s general knowledge and experience. This third level of representation is necessary because the textbase only includes propositions derived from explicitly stated information. Thus, some inferences which are needed for comprehension may not be part of the textbase. In the example in Figure 1 a drawing was used to illustrate the situation level representation of "The dog is chasing the cat". However, it should be noted that the nature or the "format" of the situation level representation is likely to depend on the experience of the individual reader (Zwaan, 2004). Zwaan and Radvansky (1998) argued that successful comprehension of a text is equivalent to constructing a coherent situation model. Thus, texts can be viewed as sets of processing instructions for the construction of a mental model of the respective situations. If such a construction is successful, a sentence is comprehended. It is important to note that therefore understanding does not only depend on the processing of external data in interaction with world knowledge, but also on the use internal information, such as beliefs, attitudes, motivations, and goals (Van Dijk & Kintsch, 1983).

Evidence for the distinction of three representational levels has been provided by several studies (Fletcher & Chrysler, 1990; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Schmalhofer & Glavanov, 1986). In an experiment by Kintsch et al. (1990, Experiment 1), participants studied sentences and, after different retention intervals, were probed for sentence recognition with a) verbatim old sentences, b) paraphrased sentences, c) inferences, and d) distractor sentences. Verbatim old sentence probes were identical to the previously studied sentences. Consequently, these verbatim old sentence probes matched the study sentences at situational, propositional, and surface level. Paraphrase sentence probes differed from study sentences at the surface level, because paraphrases involved single word or word order changes. The sentence meaning and thus the textbase and the situation model were not changed. Inference sentence probes were constructed to allow the readers to infer the respective sentence with high reliability, although surface level and textbase representations were different from the study sentences. The only correspondence for inference sentence probes occurred at the situational level. Distractor sentence probes did not correspond to the study sentences at any level of representation.

Kintsch et al.(1990) found that the recognition rate for old sentences was higher than the false alarm rate for paraphrases, which in turn were more often falsely recognized as old items than inferences. Most importantly, inferences were more often falsely recognized as old items than distractor sentences. This was taken as evidence for the influence of the situation level representation. To estimate the memory trace strengths at situational, propositional, and surface level, the authors used a signal detection theory approach (Macmillan & Creelman, 1991; Snodgrass & Corwin, 1988). They transformed the percentage of yes-responses to the different recognition probe types to *d'* values and used the following subtraction rationale. As the representations of the paraphrase sentences differed from the representations of old sentences only at the surface level, the authors proposed that the difference between the *d'* values of old sentences and paraphrase sentences reflected the strength of the surface representation. Similarly, the difference between paraphrases and inferences was used to estimate the strength of the propositional representation. Paraphrases and inferences shared the same situation level representation and they both differed with respect to the surface level representation. If paraphrases received higher endorsement rates (false yes-responses) than inferences, this difference must have originated from the propositional representation. Finally, the difference between inferences and distractors gave the strength of the situational representation by the same rationale. As inferences were different from the study sentences in the surface level and propositional representations, false positive recognition responses must have been based on the correspondences of the situation models. Subtraction of the standardized yes-responses to distractors provided a means to correct for guessing.

A key finding of Kintsch et al. (1990) was that information from the surface level and the text base decayed rapidly, whereas the situation model remained stable over a two days period. Moreover, this method of estimating representational strengths has been used to show that study goals can influence the strength of representation levels differentially (Griesel, Friese, & Schmalhofer, 2003; Schmalhofer & Glavanov, 1986) and to demonstrate that predictive inferences are only represented at the situational level (McDaniel, Schmalhofer, & Keefe, 2001; Schmalhofer, McDaniel, & Keefe, 2002).

A.1.2. Inferences

Several attempts have been made to systemize inference categories. The most commonly used classification schemes are summarized in the following. The terminology introduced here will be used to characterize sentence materials and to specify different types of inference processes throughout this thesis.
A major distinction concerns the question whether an inference is directed "backward" or "forward" (Singer & Ferreira, 1983). For illustration it is useful to consider the so-called *bridging inferences* (Haviland & Clark, 1974) and *elaborative*

inferences (Corbett & Dosher, 1978). Bridging inferences connect the current clause "backwards" to the events described in the preceding text. See the following sentence pair:

1a. *When the boy aimed at the bird with his slingshot, he hit the car´s window.*
1b. *It broke to pieces.*

To understand sentence 1b readers need to be aware of the causal relation between the two sentences. They need to "bridge" the events described in the two sentences by inferring that a hard object—presumably a stone thrown by the boy with the slingshot—broke the window. Without this bridging inference the second sentence cannot be understood. In this example the identification of a causal chain of events contributes to the coherence of the text.

An elaborative inference occurs if information is inferred which is not part of the actual text, and which per se is not necessary for understanding this text. Instead, the inference constitues an elaboration of the described situation. For example, if, after reading sentence 1a only, readers would infer that the window broke, this would be an elaborative inference. Moreover, this inference would be directed "forward", in that it predicts a future event that is likely to happen. Hence, this type of inferences has also been named *predictive inferences*.

Van den Broek, Rohleder, and Narvaez (1994) distinguished between *activation-based inferences* (associative inferences) and *coherence-based inferences*. Activation-based inferences were further subdivided into 1a) anticipation of future events, 1b) activation of preceding events, and 1c) contemporaneous information. Coherence-based inferences are assumed to be generated if comprehension is inadequate. They were classified as 2a) connecting inferences to activated information, 2b) reinstatements from long term memory, and 2c) elaborations based on background knowledge.

An even more comprehensive classification scheme was proposed by Graesser, Singer, and Trabasso (1994). The authors described thirteen classes of inferences:

1. Referential (inferences tie words or phrases to previous text elements)
2. Case structure role assignments
3. Causal antecedent (bridging inferences)
4. Superordinate goal
5. Thematic (main point or moral of the text)
6. Character emotional reaction
7. Causal consequence
8. Instantiation of noun category
9. Instrument
10. Subordinate goal/action
11. State
12. Emotion of reader
13. Author´s intent

Without going into the details of each of these different classes it becomes obvious that inferences can be made on the basis of a multitude of interactions between features of the text and internal factors of the comprehender. The focus of the current studies is on inferences of class 3) causal antecedents and class 7) causal consequences.

Another way of categorizing types of inferences takes the cognitive effort for drawing the inference into account (Graesser et al., 2007). In ascending order with increasing processing resource requirements inferences can be said to be generated: 1) automatically, 2) routinely, 3) strategically, or 4) off-line. There has been considerable debate about which inferences are actually made automatically and routinely during reading. The two major positions are expressed by the *minimalist hypothesis* (McKoon & Ratcliff, 1992) and the *constructionist theory* (Graesser et al., 1994).

According to the minimalist position only inferences which are necessary to establish local coherence, and inferences which are based on "easily available information" are encoded automatically (McKoon & Ratcliff, 1992, p. 441). Information is considered to be easily available if it has been explicitly mentioned in the text or if it is about "well-known general knowledge" (p. 441). Local coherence describes the relation of the propositions that are active in working memory at the same time (corresponding to up to two sentences). Only if local coherence cannot be achieved by the automatic encoding processes, effortful strategic and goal-based inference processes are initiated. In contrast to this minimalist hypothesis, the constructionist theory from Graesser et al (1994) assumed that readers engage in the *search after meaning principle*, and that three sets of inferences are routinely encoded: "a) inferences that address the reader´s comprehension goals, b) inferences that explain why events, actions, and states occur, and c) inferences that establish coherence in the situation model at the local and global levels" (Graesser et al., 1997, p. 182).

Graesser and his colleagues suspected that both of these models may be correct in certain conditions, depending on the goals and knowledge of the reader, and on the coherence of the text (Graesser et al., 1997, p. 183). For instance, if readers were instructed to predict the outcome of a sentence, predictive inferences were generated online (Calvo, Castillo, & Schmalhofer, 2006). On the other hand, if readers were probed for recognition memory with test words which either occurred explicitly in a study sentence or which represented an inference, predictive inferencing was not found (McKoon & Ratcliff, 1986).

In summary, these aspects are particularly important for the current studies:

- Inferences can be directed backward (bridging inferences) or forward (elaborative inferences).
- Internal factors of the reader (e.g. goals) play a role for the extent to which certain inferences are drawn.
- The minimal inferencing and the constructionist positions agree that inferences which establish local coherence (e.g., bridging inferences) are generated routinely.

A.1.3. Models of Text Comprehension

Whereas the previous section treated inference classifications and theoretical positions specific to inferencing, in the following, some broader models of text comprehension are introduced. The goal of this chapter is not to evaluate these models, but to identify and highlight important subprocesses that are needed for the generation of inferences.

Memory resonance model. In the memory resonance model (Myers & O'Brien, 1998) concepts which are currently activated in working memory serve as signals to previously processed concepts that are stored in long-term memory. Depending on the number of features shared with the input, concepts in long-tem memory "resonate", and if the resonance signal exceeds a threshold, the respective concepts are moved to working memory. This reactivation of knowledge is also referred to as *reinstatement*. An important characteristic of this model is its strict bottom-up approach to reinstatement. The resonance process is assumed to be autonomous and undirected. No active search mechanism is assumed. The model is claimed to account for several kinds of inferences, e.g. bridging inferences, causal inferences, and also predictive inferences (Cook, Limber, & Edward, 2001). In the memory-based text processing view inferences are explained as passively reactivated prior knowledge elements.

Landscape model. Similar to the resonance model, the landscape model (Van den Broek, Risden, Fletcher, & Thurlow, 1996) posits that during reading concepts in memory are activated, and these activations fluctuate. How much activation a concept receives can be defined depending on certain features of the concepts. In the example given by Van den Broek et al. (1996) the following activation levels were reported:

- Explicitly mentioned concepts result in highest activation.
- Concepts that restore anaphoric coherence, i.e. the identification of anaphoric referents, result in somewhat lower activation.
- Again lower activations are produced by two categories of concepts which establish causal coherence:
    - *Causes*, that provide sufficient explanation for an event, and
    - *Enablements*, which are only necessary preconditions for events.
- Elaborative inferences are given the lowest activation values.

In addition to the dynamic, fluctuating activation pattern that characterizes the online processing during reading, a second, more permanent memory representation of the text is built. Therefore the overall amount of activation of a concept at the end of a text is taken into account, as well as the frequency of co-occurrences of concepts. Those concepts, which are often activated together, are assumed to be conceptually related.

An interesting feature of the landscape model is it´s explicit consideration of reading goals and strategies which can differ between or within individuals. The model allows to adjust the activation weights for different reading goals and strategies—factors that are known to have an influence on inferencing (Calvo et al., 2006; Graesser et al., 1994).

Construction-Integration model. One of the most influential contribution to the text comprehension literature is the construction-integration model (CI-model; Kintsch, 1988; Kintsch, 1998). In contrast to the models presented above, the CI-model makes explicit assumptions about the structure and organization of world knowledge. The model assumes the reader´s general world knowledge is represented in a so-called knowledge net. This is an associative network in which nodes represent concepts or propositions. The nodes are connected by weighted links. Figure 2 illustrates a small part of such an associative network for the homonym BANK.

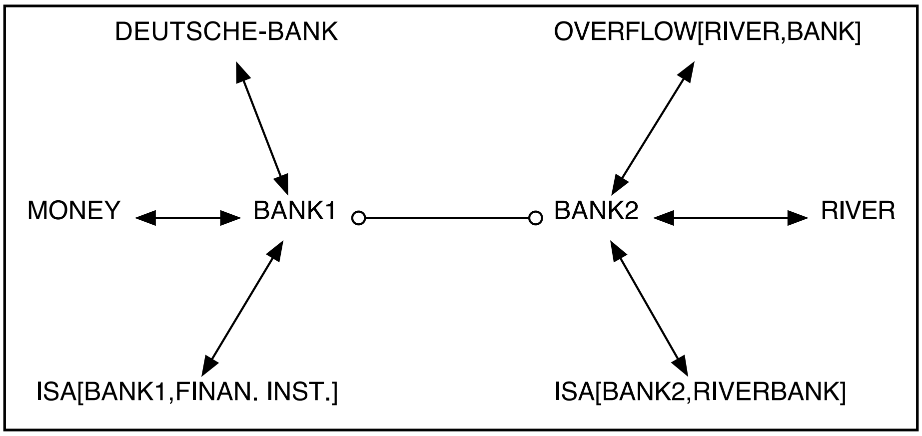


Figure 2: A fragment of the associative net for BANK (adapted from Kintsch, 1988). Positive connections are indicated by arrows, negative ones by circles.

Within this network the node BANK1, e.g., is connected to the concept MONEY and to the proposition ISA[BANK1,FINANCIAL INSTITUTE] via positive links. The nodes BANK1 and BANK2 are connected by a negative link. It can be seen that BANK1 and BANK2 derive their "meanings" from their positions in the net, i.e. from their relations to other nodes in the net. Not shown in Figure 2 are the weights associated with each link, and activation values associated with each nodes.

The CI-model proposes two main mechanisms for the construction of a coherent propositional representation of a text during comprehension: a *construction* process and an *integration* process. In the construction process a so-called enriched textbase is constructed on the basis of the input text and associated concepts from the knowledge net. The process transfers the input text into propositions. It elaborates on these by selecting closely associated neighbors, and it also infers some additional propositions from the knowledge net. The construction phase is a pure bottom-up process, and the resulting enriched textbase is often inconsistent and incoherent. In general it includes many inappropriate associates and inferences. Therefore in the integration phase these unwanted elements are excluded from the text representation. Which propositions are removed, depends on their connections to the other nodes in the network. (Technically, this is achieved by spreading activation through the network repeatedly until the activation values of the nodes stabilize.)

In summary, the CI-model proposes that during text comprehension, firstly, a lot of associations are generated in an undirected, diffuse manner, and that many of these associations are irrelevant or even dysfunctional for comprehension. Secondly, only those concepts which receive sufficient activation from interconnected neighbor concepts become part of the text representation. The CI-model assumes that inferences are represented at the situational level, as parts of the general knowledge net.

Situation Models as Simulations. Comprehension can be said to be successful if a reader was able to generate a situation model—i.e. a meaningful mental representation of the state of affairs that had been described in a text (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). Thus, the properties of situation models and the question what makes up a "good" or coherent situation model should be of central interest. Unfortunately the models described so far address these issues only peripherally (if at all). Zwaan (2004) presented a framework which focusses on situation models. It is fundamentally different from traditional approaches to language comprehension like the CI-model, in that it rejects the notion of abstract or amodal, propositional representations. The *immersed experiencer framework* (IEF) builds on theories which ground cognition in perception and action (Barsalou, 1999; Glenberg & Robertson, 2000; Pulvermuller, 1999). "The basic premise is that language is a set of cues to the comprehender to construct an experiential (perception plus action) simulation of the described situation" (Zwaan, 2004). The ongoing controversy about this approach and the possible (non-) existence and usefulness of propositional representations is not treated here (for a discussion of this topic see Singer & Leon, 2007). Importantly, the IEF shifts the research focus from the textbase to the nature of the situation model.

In the IEF comprehension is characterized by three component processes: activation, construal, and integration. During activation, words activate neuronal cell assemblies which are also involved in the perception or execution of the words' referents (so-called functional webs, Pulvermuller, 1999; Pulvermuller, 2001). Thus reading the word *fire truck* is assumed to activate a subset of neurons which are also activated during the perception of a real fire truck. The construal process integrates the activated functional webs and contrues them as a mental simulation of a specific event. Integration refers to the transition process of one construal to the next. It is assumed that attentional shifts lead to regular changes of the memory representation.

The transition process is influenced by several factors:

- concordance with human experience
- amount of overlap between the current contrual and the previous construal
- predictability, and
- linguistic cues, such word order, case markers, etc.

Inferences, according to this view, can be assumed to be constructed as part of the mental simulation of the referential situation. The notion of mental simulations that are based on previous experience is particularly important for the current studies, because recent neuroimaging research suggests that there are brain areas which contribute to similar simulation processes in several different domains. This topic will be picked up in the discussion.

A.1.4. Sentence Verification and Recognition

A variety of *online* and *memory-based methods* has been used to study language comprehension (for an overview see Zwaan & Singer, 2003). In general, online methods try to probe comprehension processes during reading or hearing. The existence of an explicit task is obligatory—participants may be instructed simply to read a text while the time they need for reading is measured. Reading times are then assumed to reflect, e.g., the processing load for different sentences. To assess the availability of information, naming and lexical decision tasks are often applied. In a naming task participants are asked to read aloud words as fast as possible. Presented probe words typically include words which occurred before in a text (or which are close associates), and words which are new. The assumption is that already activated concepts are more easily accessible which leads to faster naming latencies (priming) relative to new words. Similarly, in a lexical decision task participants need to indicate if a probe consisting of a string of letters actually is a word or not. For associates of words which were presented before, lexical decision latencies are typically shorter than for words which were not presented in the study text. Eye-tracking, event-related potentials, and PET/fMRI can also be suited for measuring online comprehension processes.
With memory-based methods a different approach is taken, which taps more into the resulting representations that are constructed during the comprehension process. Common tasks are sentence recognition and sentence verification; both of these tasks have also been used in the current studies. In recognition tasks participants are asked to answer if a test sentence has been presented before. Verification tasks usually ask participants to judge if a statement is true against the background of a study sentence or passage. The general assumption with these tasks is that they can provide insight about the strength of the memory representation of the study sentence at different levels.

So far, only the end-products of the comparison process between the representations of study sentences and test sentences have been considered. Especially in view of the use of neuroimaging methods that are based on hemodynamical activity like fMRI, it is necessary to further specify how these comparison processes could work. In other words, the process of recognizing or verifying a test sentence with respect to a previously studied sentence needs to be described in more detail. As will be shown in the following, recognition and verification both depend on the assessment of the similarity of test and study items.

Kintsch et al. (1990) proposed a process model for sentence recognition and sentence verification that is based on

- a general recognition mechanism from the list learning literature (e.g., Gillund & Shiffrin, 1984),
- the notion that discourse is represented on three levels (Van Dijk & Kintsch, 1983), and
- the processing mechanisms of the CI-model (Kintsch, 1988).

An illustration of the basic principles of this model is depicted in Figure 3. It is proposed that first the construction-integration process generates multi-level representations of the study and test sentences in form of a network of propositions. If the test sentence fits in well with the study sentence, elements of the test sentence are already part of the network of the original text. Similarity between study and test sentence can therefore be measured as the activation that flows from existing parts of the network to the nodes representing the test sentence. This similarity index is then fed into a decision mechanism. The decision component should be sensitive for various parameters, like tasks, goals, and strategies (this is not an explicit part of the original formulation of the model). Importantly, it is assumed that the comparison process takes place on all three levels of representations as soon as appropriate information is available. Naturally, during sentence comprehension, the situation model representation becomes available last.

Several other models make similar basic assumptions. Carpenter and Just (1975) state that the operations occurring during sentence verification are not specific to language processing. Basically their proposed *constituent comparison model* also describes a comparison process on the basis of two propositional representations. "To verify the sentence, the constituents in the two representations are compared—tested for identity at an abstract level. The number of operations is the primary determinant of the verification latency" (p. 47).

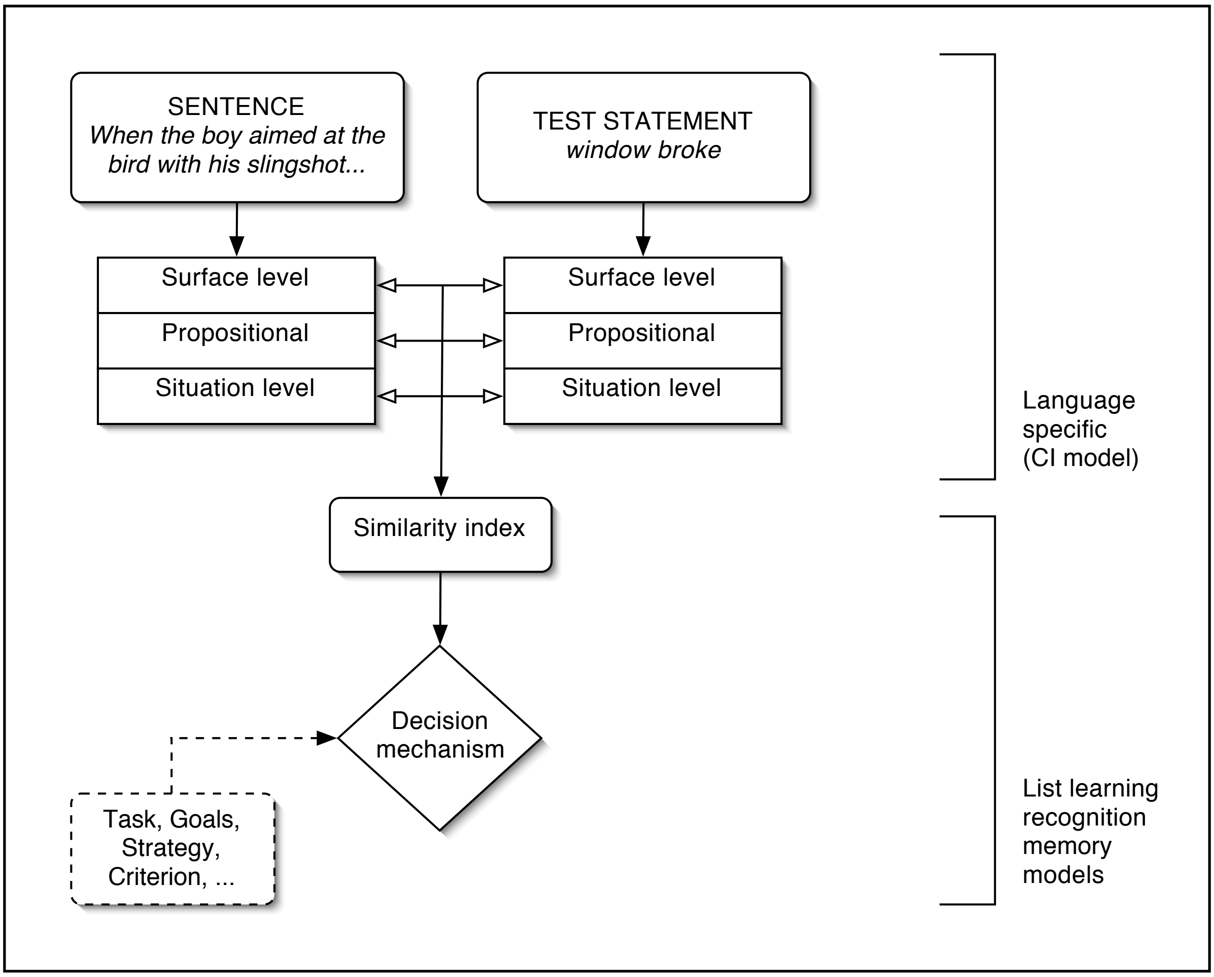


Figure 3: Sentence verification/recognition model adapted from Kintsch, Welsh, Schmalhofer, and Zimny (1990). A context sentence and a test statement are assumed to be compared on three representational levels as in the CI-model. The resulting similarity index is fed into a decision mechanism which then establishes the recognition or verification judgement.

Reder (1982) presented a multinomial two-process model for making recognition or plausibility judgements and argues that it is usually faster to judge the plausibility of a sentence than to search memory for an exact match of the test probe. Reder argues that plausibility judgements accept any number of possible facts. In contrast, direct retrieval of the test probe relies on more elaborate memory search processes because one specific piece of information must be found. Only if the representation of the test probe is highly activated in memory, recognition statements are assumed to be faster than plausibility judgements. The model proposes two stages for both recognition and plausibility judgements, representing the search process and the judgement process respectively. Kintsch et al. (1990) refer to this model and conclude that their own model explains the concept "plausibility judgement".

Lastly, the *PAM* model (Connell & Keane, 2006) presumes "concept–coherence" to be central for general plausibility judgements. Concept-coherence is viewed as consistency with prior experience. This in turn is measured as the similarity between a given scenario and prior knowledge and has led the authors to name their approach *knowledge-fitting theory*. The two-step model distinguishes a comprehension process that builds a representation of two sentences, from an assessment stage in which the plausibility is calculated. This distinction of two phases is again similar to the model of Kintsch et al. (1990) which draws on the CI-model for (1) the construction the sentence representations and (2) for determining their similarities.

A.2. Neural Indicators of Inference Processing

A number of review articles and chapters dedicated to the neural foundations of language comprehension processes have been published in recent years, reflecting the growing interest in this field (Bookheimer, 2002; Ferstl, 2007b; Gernsbacher & Kaschak, 2003; Mar, 2004). This section starts with a general introduction to the basics of functional magnetic resonance imaging. Then, after summarizing the literature on neural correlates of discourse and sentence level comprehension, mainly studies that explicitly refer to inference processes will be reviewed here. Lastly, the topic of mental simulations is picked up, and an overview of cognitive neuroscience research in this area is given.

A.2.1. Fundamentals of Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) has been used extensively to study the human brain and human cognition in recent years, and numerous introductions to the physical and statistical foundations of this technique are available to date (e.g., Cohen, 1996; Friston, 1996; Hernandez, Wager, & Jonides, 2002; Wager, Hernandez, Jonides, J., & Lindquist, 2007). Nevertheless, it might be useful to review the most important principles of so-called BOLD (Blood Oxygenation Level Dependent) imaging, as it is used in the present studies. This summary is mainly based on the literature indicated above while further readings are cited separately.

A.2.1.1. Physics

Magnetic resonance imaging (MRI) makes use of a inherent property of proton nuclei—their *spin*. Typically, the spin of a proton is illustrated as a spinning top. Spinning protons have magnetic dipole moments like submicroscopic bar magnets. Mathematically this is represented by a vector in 3-dimensional space. Brought into a magnetic field B_0, some of the spins align with the direction of the field, but—in contrast to normal bar magnets—a smaller proportion of spins also align against the direction of the field. In sum, for molecules with an odd number of protons, a net magnetization vector M results. The net magnetization depends on the strength of the magnetic field B_0. Due to the high concentration of water in the human body, hydrogen—consisting of only one proton and one electron— plays a major role for MRI for human application.

An important property of the spins is their *precession*. Like spinning tops which are tipped, the magnetization vectors precess around a vertical axis (parallel to the outer magnetic field). The velocity of this movement is called *resonance frequency* ω_0. The resonance frequency also depends on the strength of the magnetic field B_0 and, most importantly, on a constant which is specific for different nuclei—the *gyromagnetic ratio* or *Larmor frequency* γ. MRI devices are able to produce magnetic *gradient* fields with increasing or decreasing field strengths along a given axis is space. Thus, by knowing the Larmor frequency of specific nuclei in question and the magnetic field strength in a limited region in space (a slice), the resonance frequency of the specific nuclei is known.

The basic way to acquire the MRI signal is to apply a strong magnetic gradient field B_0 to a probe (e.g. the brain) and then to radiate the probe with an electromagnetic pulse B_1. This pulse (usually in the RF-range and hence named RF-pulse) is often applied perpendicularly to the main magnetic field B_0 and causes spins with a Larmor frequency equal to the frequency of the RF-pulse to flip by 90°. When the RF pulse is turned of, the spins flip back to their former orientation parallel to B_0. This local change of the magnetic field causes a detectable current in a receiving coil, which oscillates at the resonance frequency. This signal is called free *induction decay, FID*.

Let us assume that the main magnetic field B_0 is directed in the z-dimension (longitudinal) and the spins aligned with B_0 precess in the x-/y-plane (transversal). When the RF-pulse is turned off, the magnetization in the transversal plane (FID) decays and the longitudinal magnetization along with B_0 recovers. This latter process is named longitudinal *relaxation*, and it is characterized by the time constant T_1. The T_1 relaxation time strongly depends on the surrounding tissue and is also calles *spin-lattice relaxation*. Two mechanisms cause the decay of magnetization in the transversal plane. *Spin-spin relaxation* occurs because, after the RF-pulse is turned off , the spins rotating in the transversal plane get out of phase, they loose phase coherence, and consequently the net magnetization is reduced. The time constant describing this process is called T_2 relaxation rate. T_2 also depends on on the surrounding tissue. In reality, the decay of magnetization occurs much faster than T_2 because of local field inhomogeneities. This effective time constant is called T_2*.

The dephasing of spins during T_2 relaxation can be reversed by a second 180° RF-pulse, following the initial 90°-pulse after time T. In this case a second, weaker MRI signal (*spin-echo*) can be obtained after time 2T, because then the spins precess in phase again. This time is referred to as TE (time to echo). Another way to rephase the spins is by applying a gradient after the initial RF-pulse. This causes the spins to dephase faster, because the precession rate depends of the strength of the magnetic field. By reversing the polarity of the gradient the spins regain phase coherence, which then causes a *gradient-echo*. Most fMRI-studies utilize variants of the comparatively fast gradient-echo technique.

As noted above, the MRI-signal of a slice can be selected by superimposing B_0 with a gradient field in z-direction. Additionally, for reconstructing 3-dimensional images from the raw MRI signal, *frequency-encoding* and *phase-encoding* need to be applied. Frequency-encoding works by turning a gradient along, e.g., the x-axis during the measurement of the signal. The increasing strength of the magnetic field causes the spins to precess with increasing velocity in x-direction. The measured signal thus contains contributions from different frequencies which correspond to different locations in space. The reconstruction of the contributing frequencies and the corresponding locations can be achieved with *Fourier-transformation*. For phase-encoding, a series of gradients with different amplitudes is very briefly applied after the initial RF-pulse along the y-axis. These gradients cause the spins to precess with different phases in y-direction. By means of Fourier-transformation these different phases and the corresponding locations in space can be filtered out of the measured signal. With these techniques in combination, the MRI signal can be obtained in three dimensions. The elements of this 3-D space are called *voxels*.

One useful feature of MRI is the possibility to acquire signals that are sensitive to different tissues or, more general, to different molecular compositions of a probe. To a certain degree, pulse sequences, time intervals, and data collection periods can be adjusted to the specific subject of investigation. The main types of images which can be obtained are called proton-density-, T_1-, T_2-, and T_2*-weighted images. This nomenclature indicates the influence of different components of the signal to the overall image contrast. T_1-weighted images, e.g., are obtained with pulse sequences which are sensitive to T_1-relaxation.

A.2.1.2. BOLD Imaging

Functional MRI makes use of different magnetic properties of oxygenated and deoxygenated hemoglobin in blood. In deoxygenated hemoglobin the iron atoms are more exposed to the surrounding water, which causes local distortions in the magnetic B_0 field, a shorter T_2*-relaxation, and a decrease of the MRI signal (Ogawa, Lee, Kay, & Tank, 1990). Although the exact mechanisms are not fully understood, increasing local oxygen utilization during neuronal brain activity leads to an increase in blood flow and blood volume in the active region after 1 to 5 seconds. As a net result, the increasing concentration of oxygenated blood in the active region causes an increase of the MRI signal. This signal, usually acquired with T_2*-weighted imaging sequences, is called *blood oxygenation level dependent* (BOLD). The course of the BOLD signal is described by the *hemodynamic response function* (HRF). Usually, the HRF reaches it´s maximum after about 5 seconds.

It is important to note that most likely the BOLD signal only reflects a fragment of neural activity. For instance, studies with simultaneous recordings of fMRI and electrophysiological data provide evidence that "[...] activation in an area is often likely to reflect the incoming input and the local processing in a given area rather than the spiking activity" (Logothetis, 2003). Moreover, it is possible that neurons in a region are active, but the metabolic net demand of that region remains constant.

A.2.1.3. Data Preprocessing and Statistical Analysis

Before fMRI data can be analyzed, several preprocessing steps need to be executed. *Slice timing correction* algorithms are applied to adjust for the sequential acquisition of the slices. As participants usually exhibit some head motion during an fMRI experiment, *realignment* procedures correct for the displacement and rotation of the head relative to a reference image. Additionally, to better meet some of the requirements of later processing, *smoothing* of the images with a Gaussian kernel is often carried out. As the high-resolution structural images usually contain better anatomical information than the functional images, often both types of images are acquired and mapped to each other by *coregistration*. Especially when comparisons across subjects are planned, *normalization* is required to transform all images into a common space. During normalization the individual brain images are warped to match a template or a target brain.

Most statistical analysis methods are based on the *general linear model* (GLM). In the GLM, generally, one or more predictor variables are used to explain variability in a response variable. For fMRI analyses this means that one tries to build a (linear) model to predict the fMRI signal. It is assumed that the time series of signals in each voxel is composed of the additive effects of several processes (predictor variables). The contribution of each of these processes to the magnitude of the signal is unknown, but it can be estimated with the GLM. Thus, it is crucial to formulate a plausible model

In matrix notation the GLM takes the form:

$$y = X\beta + \varepsilon$$

where y denotes a column vector containing the measured data. X is the so-called *design matrix*. The design matrix contains the predictor variables in columns. β is a row vector of the predictor weights. ε is a column vector of unexplained error values. In an fMRI experiment, the β-weights can be interpreted as the magnitude of activation due to the respective predictor variables. The β-weights are found by using a least-square method that minimizes the squared distance between the observed data vector and the vector of the fitted-data, $X\beta$. If the estimated β-weights differ significantly from zero it can be concluded that the respective predictor variable is likely to contribute to the signal. Testing specific differences between conditions of an experiment, translates to building linear combinations of predictors (contrasts) in the GLM. The statistics used for testing the significance of the β-weights (often *t*-values) can then be displayed as grey-scale or color values, forming a *statistical parametric map*. Superimposing these maps on anatomical images of a brain leads to the typical illustrations of "brain activity".

Since the regression analysis is performed in every single voxel (usually of tens of thousands) the approach is a *mass-univariate* analysis. Accordingly, measures for controlling type-1 errors must be taken to avoid false positive results in many voxels. Popular methods are the application of random fields theory (for which smoothing of the data is a prerequisite) or Bonferroni correction. Also, often simply only clusters of neighboring voxels which exceed a specified "height" and spatial extent are considered. Furthermore, the analysis can be restricted to predefined regions of interest (ROI) to reduce the number of multiple comparisons.

For group comparisons, typically two separate GLM analyses are carried out. In the *summary statistics approach,* taken by software solutions like SPM2, on the first level a model is fit for each subject, and contrast images are created for each effect of interest. These contrast images are then brought into a second level group analysis, in which voxel-wise *t*-tests are performed across the contrast images of all subjects. This is often called *random effects analysis* (or *mixed-effect analysis*).

A.2.1.4. What Functional Imaging Can(´t) Tell

Since functional neuroimaging has become more and more popular for studying human cognition, it is often criticized as being a new kind of phrenology. For instance, the general doubt that imaging studies are useful for the study of cognition is expressed in Jerry Fodor's pointed remark, "If the mind happens in space at all, it happens somewhere north of the neck. What exactly turns on knowing how far north?" (Fodor, 1999).

While there appears to be broad consensus about the usefulness of mapping cognitive functions to neuroanatomical structures in the medical (neurosurgical) context, the question whether functional neuroimaging tells something about cognitive processes on a psychological level, is still a controversial issue. One of the key arguments of the proponents of functional neuroimaging is that imaging data can simply be interpreted as another dependent variable (Henson, 2005). Like behavioral measures (mostly response time and accuracy data), imaging data can inform about underlying cognitive processes and can thus constrain cognitive theories. In contrast to behavioral data, imaging data provide a direct window to the underlying brain processes. They can reveal common mechanisms underlying processes that are usually studied rather independently from each other (Jonides, Nee, & Berman, 2006). Critics of functional imaging argue that, at least in the past, imaging studies failed to provide results which can be used to distinguish between cognitive theories (Coltheart, 2006). A crucial problem, according to Page (2006), is, that cognitive theories make no predictions for neuroimaging data. Therefore imaging data cannot be used to differentiate between these theories.
In this controversy both parties invoke numerous examples for their respective positions. It seems reasonable to take a pragmatic position. Most likely, no method

can be assumed to be perfectly suited for the study of human cognition (Wager, 2006), and theories are rarely unequivocally supported by single datasets (Henson, 2006). "[...A]n understanding of the mind must emerge from a coordinated effort using converging evidence from all the tools at our disposal" (Wager, 2006). Presumably, some of the debate is caused by insufficient "coordinated effort". Functional imaging is still a relatively new method, without well-established standards, and with a lot of room for things to go wrong (Savoy, 2005). It can be expected that as the method matures, the overall quality and interpretability of functional imaging studies will increase.

A.2.2. Neuroimaging Studies on Language Comprehension

In view of the multitude of processes that are involved in language comprehension it is not surprising that many brain regions have been found to contribute to language processing. While traditionally Broca´s area in the left inferior frontal cortex, and Wernicke´s area in the left posterior temporal lobe are considered to be central "language areas", there is abundant evidence that language processing demands a larger, more complex interplay of brain regions. The "extended language network" (Ferstl, 2007a; Ferstl, Neumann, Bogler, & von Cramon, 2007) comprises regions in both hemispheres: In the temporal cortex, the anterior and mid temporal lobes (aTL, mTL) bilaterally, as well as areas all along the left superior temporal sulcus (STS) reaching posteriorly into the inferior parietal cortex (IPC; supramarginal gyrus, SMG and angular gyrus, AnG) have been found. Frontal areas include the left ventro-lateral prefrontal cortex (vlPFC) or inferior frontal gyrus (IFG), and the dorso-lateral prefrontal cortex (dlPFC). Furthermore, in several language studies bilateral activity in anterior and posterior midline-areas was found, namly in the dorso-medial prefrontal cortex (dmPFC) and the posterior cingulate cortex (PCC).

Globally ascribing specific functions to these areas is difficult for a number of reasons. First, a single region might contribute to several aspects of language processing. For example, Broca´s area, initially associated with word production, has been found to be involved in phonological, syntactical, and semantical processing—both on the word and sentence level, and both for perception and production tasks (Stowe, Haverkort, & Zwarts, 2004). To add to the potential confusion, Wernicke´s area, initially mainly associated with the comprehension of spoken language, has also been demonstrated to play a role for most of these diverse processes. How a specific region might serve different language functions is also difficult to determine and requires a detailed analysis of the respective functions. Another problem is the imprecision of localizations and labels that are often used when functions are assigned to regions. In other words, sometimes neighboring regions are subsumed under one label although they serve different functions and should be treated as different. This problem is amplified by anatomical variability and averaging techniques in group comparisons. For instance, it is often assumed that Broca´s area is located in the left IFG. This structure can be subdivided into pars orbitalis, pars triangularis, and pars opercularis (from rostral to caudal). The opercular part has been related to phonological processes and the triangular part to semantic processes (e.g., Mechelli et al., 2005).

One necessary precondition to overcome these problems is the careful analysis of the underlying processes that compose a specific language function. Large-scale analyses of language functions may also be useful to specify subprocesses. Recently, some attempts have been made to develop broader frameworks which can guide the functional anatomy of language processing. Hickok and Poeppel (2004) proposed that language processing unfolds along a *ventral stream*, which is involved in mapping sound-based representations to meaning representations in the temporal lobes, and a *dorsal stream*, which serves the mapping of sound-based representations to articulatory-based representations in frontal areas. Ullman (2004), focussing on memory systems, argued "So, whether or not there are aspects of the neurocognition of language that are unique to this faculty and to our species, much and perhaps most aspects of language are likely to not be unique" (p. 232). His *declarative/ procedural model* attempts to identify language functions that presumably evolved from phylogenetically older cognitive functions. In this model the mental lexicon depends on declarative memory structures in the temporal lobes, and mental grammar relies on brain areas which support procedural memory, such as some frontal, basal-ganglia, parietal, and cerebellar structures. An important feature of these two models is the explicit attempt to anchor language processing in a broader scope of cognitive processes.

Despite the difficulties with locating language functions, some generalizations appear to be sufficiently valid. Dominantly in the left hemisphere, anterior and posterior temporal lobe regions together with the left IFG form a core network of brain areas that is activated when language comprehension is not accompanied by an additional task. Right hemisphere homologue areas are often also activated, although to a smaller extent. This network is extended if comprehenders are confronted with an additional task (Ferstl, 2007a). The role of the anterior temporal lobes in language processing has been described as integrating text elements to meaning-based representations ("propositionalization", Ferstl et al., 2007). The posterior STS region processes phonologically-based representations (Hickok & Poeppel, 2004; Wise et al., 2001) and has been described as the region where initial semantic processing occurs (semantic activation; Jung-Beeman, 2005). Regions in the left temporal lobe are engaged in storing conceptual knowledge (Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Indefrey & Levelt, 2004; Jung-Beeman, 2005). The left IFG has to be segregated into subsystems. While the anterior part (pars triangularis) seems to be involved in semantical processing (Gitelman, Nobre, Sonty, Parrish, & Mesulam, 2005), the posterior part (pars opercularis) seems to be related to phonological processing (Poldrack et al., 1999).

An ongoing controversy exists whether the right hemisphere plays a particular role in discourse comprehension. Beeman proposed that the right hemisphere is engaged in so-called *coarse semantic coding* (1993; Jung-Beeman, 2005). The key idea of this account is that the activation of semantic concepts in the right hemisphere occurs in a less fine-tuned and selective manner than in the left hemisphere. Thus, activation spreads from one concept to more distantly related concepts, and this more diffuse activation spread helps to detect conceptual relations between sentences and supports inferencing. The framework is based on two principles:

1. Three semantic processes build the core language comprehension network:
    a. semantic *activation* in the pMTG
    b. semantic *integration* in the aTL
    c. semantic *selection* in the IFG.

2. Each semantic process occurs in both hemispheres, but the RH performs coarser semantic coding, relative to the finer semantic coding in the LH (Jung-Beeman, 2005)

While initial evidence for this approach was collected with patient-studies, visual-hemifield presentation techniques, and ERP-studies (Beeman, 1993; Beeman, 1998; Jung-Beeman, Bowden, & Gernbacher, 2000; Kiefer, Weisbrod, Kern, Maier, & Spitzer, 1998), some results from fMRI-studies were also interpreted as to support this proposal (Mason & Just, 2004; Virtue, Haberman, Clancy, Parrish, & Jung Beeman, 2006). However, Ferstl et al. (2007), in a meta-analysis of 23 neuroimaging studies on text comprehension, concluded that no reliable evidence for a special role of the right hemisphere for inference processing could be established.

A.2.3. Neuroimaging Studies on Inference Processing

There are some neuroimaging studies that refer directly to inference processes in the brain. So far, the predominant paradigm has been to manipulate the coherence between short sentences, because different theoretical accounts on inferencing agree that inferences are generated automatically if they are needed to establish local coherence between subsequent sentences (Graesser et al., 1994; McKoon & Ratcliff, 1992).

In an fMRI-study on causal inferences Mason and Just (2004) varied the causal relatedness of two-sentence passages. Four conditions were realized by varying the content of the first sentence while the second sentence was constant over conditions:

1.	Highly related:	*Joey´s brother punched him again and again.*
2.	Moderately related 1:	*Racing down the hill, Joey fell off his bike.*
3.	Moderately related 2:	*Joey´s crazy mother became furiously angry with him*
4.	Distantly related:	*Joey went to a neighbor´s house to play.*
→	"Outcome sentence":	*The next day his body was covered with bruises.*

Cortical activity was measured during the reading of the passages in three large regions of interest:

– a left hemisphere language network comprising IFG, TL, and IPL,
– the homologue areas in the right hemisphere language network, and
– the bilateral dorso-lateral prefrontal cortices.

The authors predicted different patterns of activations in these networks depending of their function for the construction and integration of inferences based on the CI-model (Kintsch, 1988). Regions supporting inference generation should show increasing activity with decreasing conceptual relatedness. Integration processes, on the other hand, should be characterized by maximal activations for moderately related sentences. Consistent with these assumptions, the results indicated that the bilateral dorso-lateral prefrontal cortices (dlPFC) might be involved in the generation of inferences, and that right hemisphere language areas seem to play a role in the subsequent integration of inferences into the reader´s text representation.

Identifying some methodological problems in the aforementioned study, Kuperberg et al. (2006) conducted an fMRI study with similar sentence materials and an additional coherence judgement task. Their texts consisted of two context sentences and a target sentence:

1. Highly related: *The boys were having an argument.*
   They began hitting each other.

2. Intermediately related: *The boys were having an argument.*
   They became more and more angry.

3. Unrelated: *The boys were unsure about the weather.*
   At noon they started to hike.

→ "Target sentence": *The next day they had bruises.*

Like the authors assumed, intermediately related scenarios induced more inferencing activity than either highly related or unrelated scenarios. A bilaterally distributed network of areas was identified that mediated causal inferences, including lateral and medial portions of the prefrontal cortex, the middle and superior temporal gyri, as well as the inferior parietal lobes. Thus, the study provided support for the notion that inference processing is accomplished by areas in the extended language network.

Ferstl and von Cramon (2001) also studied the coherence of sentence-pairs using fMRI. Their participants had to decide whether two visually presented sentences were pragmatically related, and the passages were either coherent, or the sentences were incoherent:

Coherent: *Sometimes a big truck drives by the house. The dishes start to rattle.*
Incoherent: *The lights have been on since last night. The dishes start to rattle.*

Contrary to studies that report right hemisphere activations, the authors found two areas—one in the left dorsomedial prefrontal cortex (dmPFC) and one smaller region in the posterior cingulate gyrus and precuneal area—that showed higher activations in coherent trials than in incoherent trials. The authors related the activation in the medial portion of the superior frontal gyrus to the initiation of inference processes. More generally, they claim that this region is activated when a task requires cognitive processes that depend on the utilization of individual background knowledge. These results were furthermore replicated with a corresponding listening comprehension task (Ferstl & von Cramon, 2002). Additionally, this study demonstrated that the dmPFC is important for the initiation of inference processes regardless of the animacy of the sentence referents. This is important since the dmPFC has also been found to be involved in *theory-of-mind processing*, which is described as the human "ability to explain and predict other people's behaviour by attributing to them independent mental states" (Gallagher & Frith, 2003).

Another experiment by Ferstl and her collegues (Ferstl, Rinck, & von Cramon, 2005) studied emotional and temporal aspects of situation model processing in short stories. They assumed that inconsistencies in the emotional experience of the protagonists would initiate explanatory inferencing, whereas chronological inconsistencies in the chain of story events would trigger a reinstantiation search and comparison process in the memory for the previous discourse. If, for instance, a character in the story expresses sadness although earlier in the story the reader learned that he or she was quite happy at a cheerful party, a causal explanation for a potential change of feelings is looked for. A similar explanation is not possible for the kinds of chronological inconsistencies which were implemented because these were all or none. If it is mentioned in the text that Markus arrived earlier than Claudia, it is inconsistent if later on it was stated that Claudia already waited for Markus. The participants' task was to indicate whether they detected inconsistencies after the reading of each story. In line with the expectations, the dmPFC was found to respond to inconstistent emotional stories more than to consistent emotional stories. Both, emotional and chronological inconsistencies evoked higher activity in the bilateral vlPFC than their consistent counterparts. For chronological inconsistencies this effect was enhanced, which was explained with the higher likelihood of reinstatement processes for this condition.

Virtue et al. (2006) conducted an fMRI-study in which participants read stories that consisted of multiple sentences without a concomitant secondary task. An example story started with this passage:

> *Nancy called John to say she'd pick him up early for her best friend's wedding. John had been sitting around the house in his jeans, so he went to his bedroom to find*[1] *some clothes/to change his clothes. Soon he came out wearing his tuxedo*[2] [...]

The authors measured neural activity at two critical points in the text—marked by superscripted numbers in the example passage—namely when a verb implied an inference or mentioned an action explicitly ([1]verb point) and when an inference was needed to establish coherence between successive sentences ([2]coherence break). Two regions were predicted to be more activated in the implication condition than in the explicit condition. The authors assumed that the STG is involved in the semantic integration of inferred information. Greater IFG activity was predicted because semantic selection of inferential information should be supported by this region. Moreover, it was assumed that this pattern should be more pronounced for participants with high working-memory capacity as compared to participants with low working-memory capacity. The results showed that inference generation was associated with activity in the right superior temporal gyrus at the verb point, whereas at the coherence break, activity was predominantly found in the left superior temporal gyrus. The high memory capacity group additionally showed higher activation in the left IFG at the coherence break.

Directly comparing the results from the aforementioned studies is difficult for several reasons. For example, most studies differ considerably with respect to experimental paradigm and imaging data analysis. Table 1 gives an overview of some methodological aspects concerning the tasks, materials, and data analyses used. A variety of tasks were assigned to the participants, including reading for comprehension without concomitant task (although additional comprehension questions were included in non-scanned periods of the experiments), rating the difficulty to connect sentences, judging whether sentences are related, and detecting inconsistencies. Predominantly, texts were presented visually; two studies used auditory presentation. Text materials ranged from two sentence passages to short stories, and the experimental manipulations aimed at causal coherence, cohesion, consistency with respect to emotions or time, and implicit or explicit mentioning of facts.

A corresponding diversity exists for the imaging data analyses, in particular with regard to the structure of the chosen analysis models. Activations were estimated defining either events which were time-locked to critical words, or epochs of different length were defined. In case of the existence of an overt tasks, the participants' responses were either included in the epoch of interest or they were excluded. Except for one case (Mason & Just, 2004), in all studies almost the entire brain was scanned, and regions of interest analyses supplemented the whole brain analyses.

In view of the methodological differences across studies, varying results and sometimes diverging conclusions are not surprising. Most notably, each set of text materials together with the associated task instructions is likely to trigger more or less different processing strategies—potentially leading to different activation patterns.

Table 1: Imaging studies on inferencing.

Study	Task and materials	Analysis
Mason & Just (2004)	– Reading of sentence pairs varying in causal relatedness – Comprehension probes in filler trials	– Large, pre-defined ROIs, dmPFC and anterior TL not covered – ANOVA on number of activated voxels (height threshold) – 22 sec epochs (10 sec sentence reading + 12 sec rest)
Kuperberg et al. (2006)	– Rating the difficulty of connecting a visually presented target sentence with a 2-sentence passage – Manipulation of causal relatedness	– HRF-model with two epochs: a) trial onset to critical word (9.33 sec), and b) rest of the trial including response (4.66 sec) – FIR-model in ROIs
Ferstl & von Cramon (2001)	– Yes-/No-judgement if two visually presented sentences are pragmatically related – Manipulation of coherence and cohesion	Event of interest time-locked to the onset of the last part of the target sentence
Ferstl & von Cramon (2002)	– Manipulation of coherence with two types of auditorily presented materials: A) sentences without animate agents, and B) sentences with reference to people – Task for A: Yes-/No-judgement if sentences are logically connected – Task for B: Yes-/No-judgement participants can identify with the people in the passage	Event of interest time-locked to 1 sec before offset of the target sentence
Ferstl et al. (2005)	– Detection of inconsistencies in auditorily presented passages of seven sentences – Stories were either consistent/ inconsistent with respect to emotions or time	– Event of interest time-locked to offset of the critical word – Epoch of interest spanning the time between target word and story ending
Virtue et al. (2006)	– Reading of short stories – Answering true/false questions between scanning runs – Manipulation of verbs which either implied an inference or stated a fact explicitly	Epochs of 6 sec time-locked to critical words (verb point and coherence break)

Therefore, in addition to careful evaluation of the tasks and analyses at hand, replications of findings are of particular importance. Overall, the studies reviewed indicate that inference processing most likely takes place with important contributions of the dmPFC and/or regions in the right hemisphere. In the following the robustness of these previous findings is discussed in more depth.

The only study directly replicating conditions of another experiment with only minor modifications is Ferstl and von Cramon (2002). These authors confirmed their previous finding that the dmPFC might be related to the establishment of coherence (Ferstl & von Cramon, 2001). The involvement of the dmPFC in inferencing was also suggested by Ferstl et al. (2005) and Kuperberg et al. (2006), and hence seems to be a relatively robust finding. On closer inspection it stands out that in two of these studies coherent sentences elicited *higher* activation than incoherent sentences (Ferstl & von Cramon, 2001; Ferstl & von Cramon, 2002), whereas in the other two studies coherent texts partially elicited *lower* activation than less coherent texts (Ferstl et al., 2005; Kuperberg et al., 2006). A possible explanation for this apparently contradictory finding might be found in the materials and tasks utilized. In Ferstl and von Cramon (2001; 2002) the range of causal relationships between the sentence pairs was rather limited—i. e., the sentences were either related or they were not (see also Kuperberg et al., 2006, for a discussion of this topic). Therefore, coherence building processes may have been more intense in the coherent condition since incoherent sentence pairs were easily identified as such. This was not the case in Kuperberg et al. (2006) who used three response categories (closely, intermediately, and distantly related), and whose “unrelated” scenarios still could be causally connected with some effort. In addition, comparing the text examples of these studies reveals that most probably the intermediately related scenarios roughly correspond to the coherent sentence pairs of Ferstl and von Cramon (2001; 2002). Thus, concerning the respective conditions which were most likely to induce inference processing in the different studies the results are compatible and point to an involvement of the dmPFC. Similarly, the nature of the longer text materials used by Ferstl et al. (2005) may have induced inference processing to a larger extent when inconsistencies occurred, as compared to consistent stories.

Two studies particularly highlighted the role of right hemisphere areas for inference processing. It is noteworthy that these studies did not demand additional, overt tasks apart from comprehension during imaging data acquisition. Virtue et al. (2006) associated the right STG with the generation of inferences. Mason and Just (2004) proposed that the integration of inferences into the discourse representation is achieved with contributions of a right hemisphere language network. Unfortunately, Mason and Just (2004) only reported activations that were averaged across quite large regions of interest. It is therefore not possible to evaluate if there was overlap with the activation in the STG found by Virtue et al. (2006). Another piece of evidence supporting the notion of right hemisphere involvement in inference generation comes from Ferstl et al. (2005) who found the right anterior temporal lobe to be more strongly activated in response to inconsistent information than to consistent information. This activation occurred in more anterior parts of the temporal lobes compared to the activations reported by Virtue et al. (2006). Ferstl et al. (2005) interpret their results concerning the right anterior temporal lobe in terms of a general supporting function of right hemisphere areas. In this view right hemisphere areas do not subserve functionally distinct processes from left hemisphere areas. Instead they assist the functions of the left hemisphere in case of increased processing difficulty.

Overall there seems to be less consistent evidence for the right hemisphere view on inference processing with respect to functional neuroimaging studies. Ferstl (2007b) speculated that reasons for these varying results might be the larger anatomical variability in the right hemisphere across participants, or that the right hemisphere supports qualitatively different processes which are not quite understood yet. As noted above, the imaging studies primarily advocating that inferencing takes place in the right hemisphere (Mason & Just, 2004; Virtue et al., 2006) measured brain activity in absence of additional tasks during data acquisition. Although clearly more evidence is needed to draw conclusions from this observation, it suggests that qualitatively different subprocesses may be more or less prominent during different tasks.

A further aspect of the previous research on inference processing is noteworthy. Due to the structure of the text materials used in most studies inference processing and general situation model building processes are confounded (Ferstl, 2007b). Consider again the 'coherent' sentence pair example of Ferstl and von Cramon (2001): *Sometimes a big truck drives by the house. The dishes start to rattle*. The cognitive processes needed to comprehend this passage not only involve the generation of the inference that vibrations produced by the truck are the cause of the rattling dishes, but also the generation of a situation model representation of the rattling dishes themselves. This problem also affects the interpretation of the results of Ferstl and von Cramon (2002), Kuperberg et al. (2006), as well as Mason and Just (2004). In these studies changes in coherence are accompanied by general changes of the situation model representations. The current studies, presented in the following, avoided this confound by using text materials which were constructed to keep general situation model building processes constant across the most important experimental conditions.

A.3. The Present Experiments

Obviously, the picture that neuroimaging research on language comprehension has generated so far is still incomplete. The specific functions of the regions in the extended language network are in part unknown, and how these regions cooperate to implement inference processing is unclear as well. On the other hand, considerable progress has been made in identifying brain areas which are likely to support inference processing, and attention has been directed to the possible influence of task differences and text materials.

In this research a well-established theoretical framework of text comprehension (Van Dijk & Kintsch, 1983) was used to be able to identify important components of inference processes. The focus was put on verification and recognition processes, because the verification of text ideas can be assumed to be central to general language comprehension and inferencing (Singer, 2006). Moreover, verification and recognition processes have been shown to be closely related (Kintsch et al., 1990). As has been outlined in A.1.4, verification and recognition both depend crucially on comparisons of text representations, and the text comprehension model of van Dijk and Kintsch (1983) offers a systematic approach to the investigation of the processes tapping into these representations during language comprehension. This theoretical background provided the basis for a further specification of the functional neuroanatomy of inferencing with respect to verification and recognition processes.

The overall strategy of the two experiments was two-fold: First, to isolate inference processes from general situation model building processes, sentence material was used which allowed to assess processes on the different representation levels. Second, different task instructions were realized in the two experiments to selectively intensify situation level processing (verification instruction in Experiment 1) and text level processing (recognition instruction in Experiment 2). It was expected that in both experiments regions within the extended language network would be activated by the respective tasks, and that the combination of task instruction and processing level discrimination would allow to identify the functional contributions of the involved brain regions.

For the experiments sentence materials were translated and modified which had been used in numerous previous behavioral studies on inference processing (e.g. Potts, Keenan, & Golding, 1988; Keefe & McDaniel, 1993; McDaniel et al., 2001; McKoon & Ratcliff, 1986; Schmalhofer et al., 2002). These materials consist of sentence pairs like *While the flight attendant served the passenger a glass of red wine turbulences caused the wine to spill*. Referring to this context sentence, the statement *The wine was spilled* can be characterized as an explicit repetition. By modifying the ending of the context sentence, different relationships between the representation of the test statement *The wine was spilled* and the representation of the context sentence can be established. If the context sentence of the example above ends with a paraphrase of the verb "spilled": *turbulences caused the wine to splash*, the test statement´s representation differs from the explicit version mainly in the surface representation. In this case, the representations at the situational level and the propositional level overlap to a large degree. The "inference" ending *turbulences occurred which were very severe* provides enough information for the reader to infer, that *The wine was spilled* is a plausible consequence. With regard to the context sentence, McDaniel et al. (2001) have shown, that the inference (that the wine was spilled) is only represented at the situational level. Finally, if the context sentence is "unrelated" in that it does not provide an explanation for the test statement (*While the flight attendant served the passenger a glass of red wine, the plane was at cruising altitude*), it can be assumed that the statement is not part of the representation of the context sentence at all. See Table 2 for an example of the original German material together with an English translation.

Table 2: Experimental conditions and example text material. For the four main conditions (explicit, paraphrase, inference, and unrelated) variations of a common theme were created, consisting of a headline, an 18-words passage, and a test statement. Additionally, filler materials and pseudoword sequences were constructed.

Context sentence	Test probe
Explicit condition	
Die Steinschleuder Als der Junge auf den Vogel zielte, traf er dabei das Fenster des Autos, s*o dass es zerbrach*.	„Fenster zerbrochen"
The slingshot When the boy aimed at the bird, he hit the car´s window *which broke.*	"window broke"
Paraphrase condition	
Die Steinschleuder Als der Junge auf den Vogel zielte, traf er dabei das Fenster des Autos, *so dass es zersplitterte.*	„Fenster zerbrochen"
The slingshot When the boy aimed at the bird, he hit the car´s window *which splintered.*	"window broke"
Inference condition	
Die Steinschleuder Als der Junge auf den Vogel zielte, traf er dabei das Fenster des Autos *mit einem großen Stein.*	„Fenster zerbrochen"
The slingshot When the boy aimed at the bird he hit the car´s window *with a big stone.*	"window broke"
Unrelated condition	
Die Steinschleuder Als der Junge auf den Vogel zielte, *wurde seine Mutter sehr wütend und nahm ihm die Steinschleuder weg.*	„Fenster zerbrochen"
The slingshot When the boy aimed at the bird, *his mother became angry and took away the slingshot.*	"window broke"

Table 2 (cont.)

Context sentence	Test probe
Filler material	
Das Fernsehprogramm Christian schaltete von einem Fernsehprogramm zum Nächsten aber nichts wollte ihm gefallen. Also machte er den Fernseher aus.	Fernseher repariert"
The television program Christian zapped from one channel to the next, but nothing pleased him. He switched off the TV.	"Television repaired"
Pseudoword condition	
Pseudoworte / Pseudowords Uds wur fobel gumdo qij lev kujs dorbig anbyv wafsuv falnul doj oenim woo mo kohnro seepu jubdo	„seepu jubdo" / „uds wur"

Based on previous work (Kintsch et al., 1990; McDaniel et al., 2001; Van Dijk & Kintsch, 1983), it can be argued that the representations of explicit test statements match those of the context sentences at surface, propositional, and situational level. Paraphrase statements share corresponding representations only at the propositional and situational level. In contrast, for inference statements solely the situation level representations match. Lastly, the representational levels of unrelated statements cannot be mapped to the representation of the context sentence at all.

Consequently, these differences in the representations typically lead to differential behavioral response patterns in sentence recognition tasks. Usually, explicit sentences are recognized with the shortest response latencies, followed by paraphrase sentences, inferences, and unrelated sentences (e.g. Schmalhofer & Glavanov, 1986). Correspondingly, the proportion of "yes"-responses decreases in the same sequence—it is highest for explicit sentences and lowest for unrelated sentences.

The differences between these endorsement rates have been used to estimate the strength of the memory trace of the test sentence at the three representational levels in the following manner (Kintsch et al., 1990). The difference between explicit and paraphrase conditions is supposed to reflect the strength of the surface level representation. An estimate of the propositional representation is given by the difference between paraphrases and inferences, and finally the strength of the situational representation can be approximated by the difference between inferences and unrelated sentences. In the current studies, we used an analogous rationale to define contrasts that characterize the neural correlates of processing at the different levels of representation.

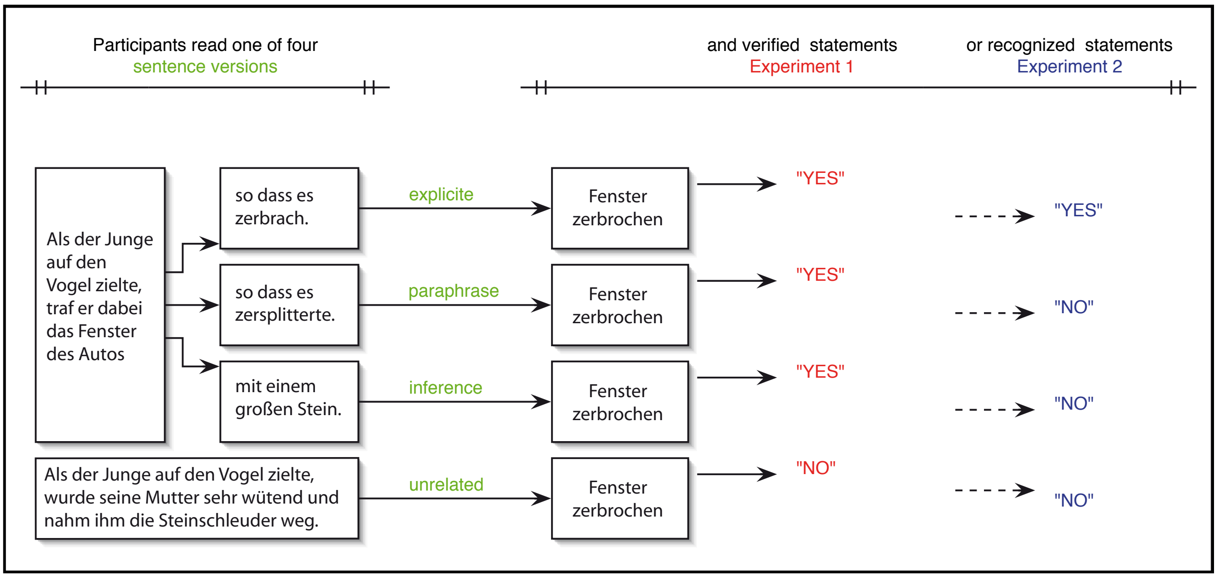


Figure 4: Overview of experimental conditions and tasks in Experiment 1 and Experiment 2. See Table 2 for an English translation of the example sentence. In both studies participants read one of four sentence versions (explicit, paraphrase, inference, or unrelated/control). A test statement subsequently presented had to be verified in Experiment 1 and recognized in Experiment 2. The verification task asked if the statement was sensible in the context of the previously read sentence. In the recognition task it was asked if the test words appeared in the context sentence. Participants were required to respond with a yes- or no-button. An additional pseudoword condition and filler materials are not illustrated in the figure (see the text for more details).

With respect to the comparison processes that work at the different levels of representation the following assumptions were made:

1. It was supposed that both the context sentence and the test statement can be accessed in working memory (Graesser et al., 1997). In the explicit condition, the surface-level match of verification statements and the corresponding words in the context sentences can therefore be detected within the working memory system.
2. For comparisons at the propositional level semantic memory has to be tapped.
3. Determining the relations between situational representations additionally requires more constructive processes that are closely linked to the episodic memory system (Kintsch et al., 1990; Lea, Mulligan, & Walton, 2005).

Figure 4 provides an overview of the experimental conditions in the two experiments. The participants in the two studies read sentences and immediately afterwards decided if a short target statement was true with respect to the sentence just read (verification task in Experiment 1), or they decided if the statement was presented before (recognition task in Experiment 2). Hemodynamic cortical activity was measured with fMRI BOLD imaging during the reading phase and during the execution of the verification or recognition task.

B. Experiment 1: Verification

In Experiment 1[1], a verification task was used to enhance inference processing since processing goals have been shown to influence the amount of inferences comprehenders draw during reading (Calvo et al., 2006; McDaniel et al., 2001). For example, Griesel et al. (2003) used similar sentence material and found that participants made three times as many bridging inferences with a verification instruction as with a recognition instruction (the proportions of yes-responses were .81 for verification versus .27 for recognition).

Four main types of pairings between context sentences and verification test statements were used. With respect to the context sentence the verification probes were classified as explicit, paraphrase, inference, or unrelated statements (Table 2).

The key assumption of Experiment 1 was that readers needed to generate an inference to connect sentence pairs like the following:

(1) *When the boy aimed at the bird he hit the car´s window with a big stone*
(2) *The window broke.*

More precisely, for connecting the two sentences in a coherent manner, a bridging inference was to be drawn. Different theoretical positions agree that this type of inferences is generated online during normal reading (Graesser et al., 1994; McKoon & Ratcliff, 1992). Less agreement exists concerning the question if a predictive inference is already constructed after the reading of sentence (1). However, several researchers have provided evidence for the notion that the inference, represented by sentence (2), is represented at the situational level (Fincher-Kiefer, 1995; Fincher-Kiefer, 1996; McDaniel et al., 2001; Schmalhofer et al., 2002).

Thus, a verification decision in the case of this type of inference sentence pairs is primarily based on a comparison of the situation level representations of the two sentences. Although the situation level representation is crucial for the verification, it is assumed that comparison processes also take place at the propositional level as well as at the surface level representations (Kintsch et al., 1990). Therefore, neural activity measured during the processing of inference sentence pairs would reflect processing at all three levels of representation. Brain areas which are primarily involved in inference processes on the situation level can be identified by contrasting (subtracting) the processing of inference statements and paraphrase statements. For paraphrase statements the verification process can be based primarily on a match of the propositional representations of context sentence and test statement. Extensive comparisons on the situational level are not necessary. As in both, inference and paraphrase condition, the surface level representations of context sentence and test statements differ, the processing at this level can be assumed to be factored out.

With the same rationale, processing at the propositional level can be assumed to be tapped by the contrast of paraphrase and explicit statements. In these conditions

1 The study reported here is identical to Experiment 2 of Friese, Rutschmann, Raabe, & Schmalhofer (in press) .

context sentence and test statement share the same textbase but differ with respect to the exact wording. Therefore, more cortical activity in response to paraphrase statements than to explicit statements should reflect the evaluation of the sentence meaning at the propositional level.

Based on the review of the neuroimaging literature on inference processing it can be expected that the reading of the context sentence and the verification of the test statements are supported by regions in the extended language network. Specifically, candidate areas for inference processing at the situational level—as it is presumably revealed by the contrast of inference and paraphrase conditions—include the right hemisphere (Jung-Beeman, 2005), the bilateral dlPFC (Mason & Just, 2004), and the dmPFC (Ferstl & von Cramon, 2001). Textbase-level processing should activate areas in the left temporal lobe (Indefrey & Levelt, 2004).

B.1. Methods

B.1.1. Design and Material

A total of 72 text scenarios was constructed in German. Each scenario consisted of a headline, four sentence variations and one test statement. With reference to the test statement the sentence variations constituted (a) an explicit trial, (b) a paraphrase trial, (c) an inference trial, or (d) an unrelated trial (Table 2). The general structure of all sentences was that a situation involving some actions was described. Furthermore, the outcome of this action was mentioned either explicitly or implicitly in the following way. (a) Explicit sentences contained the same words as the test statements. In (b) paraphrases, synonymous phrases were used. The (c) inference sentences did not mention the outcome in question directly, but provided sufficient information to infer it. Lastly, the (d) unrelated sentences provided only insufficient and incoherent information regarding the test statement as a possible outcome of the situation described.

Each sentence consisted of eighteen words. The first 13 (± 1) words were identical in the explicit, paraphrase, and inference condition. To assure that each sentence appeared equally often in all conditions across participants, the sentence pool was divided into four lists, and lists and conditions were counterbalanced by a latin square. The expected yes- and no-responses were balanced by including additional eighteen filler sentences that were supposed to be correctly answered with "no" like the unrelated sentences. Furthermore, eighteen pronounceable pseudoword sequences were constructed by permuting individual letters in sentences which were not contained in any of the other experimental conditions. This pseudoword condition was included as a baseline task for the fMRI-study. For the pseudoword sequences the verification statement consisted of the first two or the last two pseudowords of the sequence. The statement had to be answered with "yes" when the pseudowords in the verification task statement were the same as the last words of the sequence. This baseline task is similar to the verification task in that it includes basic memory and decision processes, but lacks language specific coherence building processes. The presentation of the total 108 trials was pseudo-randomized to provide three blocks of 36 trials, six of each of the five experimental conditions (pseudowords, explicit, paraphrase, inference, unrelated) plus six filler trials.

B.1.2. Procedure

All participants were informed about the type of stimuli presented to them and instructed to give a positive response only if the verification statement was sensible with respect to the preceding sentence or, in case of pseudoword sequences, if the two items of the test probe were identical to the last two items presented in the pseudoword sequence before. They were also asked to read the pseudoword sequences as if it was normal text.

Stimuli were generated with E-prime (Psychology Software Tools, Inc., Pittsburgh, USA) and presented on a 22" CRT screen. Responses were recorded with a serial response box (Psychology Software Tools, Inc., Pittsburgh, USA) which the participants operated with their right hand (index finger for “yes”, middle finger for “no”).

The context sentences were presented word-by-word, centered on the screen (rapid serial visual presentation) with 600 ms presentation time for the headline, followed by a blank screen for 600 ms and then successively 300 ms for each of the 18 words with an interstimulus interval of 300 ms (12 seconds in total). After the last word a central fixation cross appeared for two seconds, followed by the test statement for a maximum of 4.3 seconds. The test statement disappeared from the screen after the participants´ responses. After the response time window of 4.3 seconds a central fixation cross was presented for a pseudo-randomly varied interval of 7-11 seconds. In total, one trial lasted for 27 seconds (see also Figure 5).

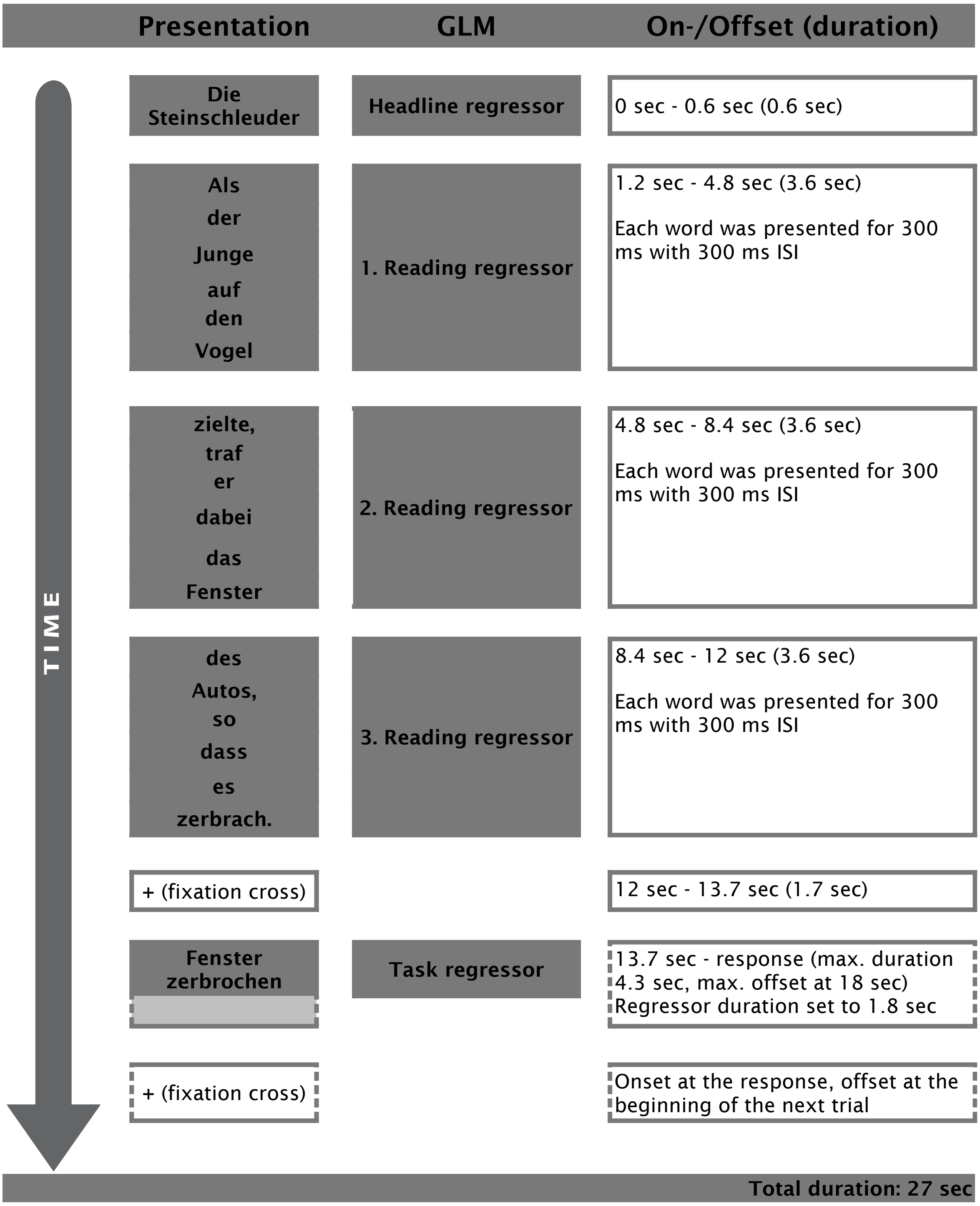


Figure 5: Temporal structure of a trial and relation to GLM regressors. Each word was presented serially for 300 ms. Afterwards, the test statement was presented as a whole until the participant´s response. For the GLM-based analysis the reading phase was modelled as three episodes referring to six words each. The presentation of the test statement and the response were modelled as a separate episode of 1.8 seconds length.

The experiment was preceded by seven training trials. Subsequently, a total of 108 trials followed (18 trials each for explicit, paraphrase, inference, unrelated, filler, and pseudoword condition). After blocks of 36 trials short resting periods of up to three minutes were given upon demand of the participants. The whole experimental procedure lasted for approximately 65 minutes.

B.1.3. Participants

Thirteen participants volunteered in the fMRI study for course credit or payment. Seven participants were female, and the average age was 22.8 years. All participants were right-handed, native speakers of German. They were healthy, had no history of neurological illness, and all had normal or corrected to normal vision (contact lenses). Informed written consent for participation in the fMRI study was obtained from all participants.

B.1.4. fMRI Image Acquisition

MR-images were acquired at the MR-facility of the University of Oldenburg in the lab of Prof. Mark Greenlee with a 1.5 T Siemens Sonata whole body MRT equipped with an 8-channel head array coil (MRI-Devices Europe Inc., Würzburg, Germany). Head fixation was achieved by using soft pads. Foam ear plugs and sound damping headphones were used for noise shielding.

During the functional scans the blood oxygen level dependent (BOLD) response was measured using a T2*-weighted gradient EPI sequence (TR = 3 s, TE = 50 ms, flip angle = 90°, resolution 3 x 3 mm^2, number of slices = 35, interleaved acquisition sequence, slice thickness = 3 mm, distance factor: 0-10 %). The acquired slices were rotated approximately 10° relative to the AC-PC line in order to cover prefrontal, parietal, and temporal regions in full and the majority of the occipital cortex, sometimes excluding the ventral extent of V1. A total of three functional sessions of 326 images each was recorded. The onset of the stimulus presentation was jittered with 0, 1, or 2 seconds relative to the onset of the image acquisition to reach a virtual time resolution of 1 second.

Structural images were acquired for each participant using a T1 weighted MPRAGE sequence (TR = 1900 ms; TE = 3.93 ms; resolution 1 x 1 x 1 mm^3) at the end of the experiment.

B.1.5. fMRI Preprocessing and Modelling

All MR-data were preprocessed and statistically analyzed using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). Functional images were corrected for acquisition delays and realigned to the first image. A mean image of the functional images was computed, and the structural image was coregistered to the mean of the functional images. After normalizing all images to the MNI-152 template, realigned functional images were resampled to 2 x 2 x 2 mm^3 and spatially smoothed using an isotropic Gaussian kernel with 10 mm FWHM.

Two types of statistical analyses were performed—one based on the whole brain volume and a second on functionally defined regions of interest (ROI). The ROI analysis was included to investigate the effects of the experimental manipulation in regions known to be important for language processing with more statistical power. For both analyses a general linear model was fitted to each individual data set, modeling the presentation of the headline, the subsequent sentence and the verification probe for each condition separately. The modeling of the sentence presentation was split into three regressors corresponding to six words each. The verification process was modeled using a block with a duration of 1.8 seconds starting from the onset of the test statement. This duration was chosen based on the participants´ largest average response time to inference statements. Individual t-test contrasts were calculated between verification tasks in the inference, explicit, paraphrase, and unrelated conditions as well as for the pseudoword condition. For statistical analysis a random-effects model was used to test for the presence of significant activation clusters (*t*-statistics). Additionally the time courses of activation were extracted from the regions of interest to reveal any differences in the shapes of the hemodynamic response functions.

B.1.5.1. Whole brain analysis

For the whole brain analysis, statistical maps were thresholded with an uncorrected p-value of $p < .001$ ($t = 3.93$). Unless otherwise noted only clusters surpassing a corrected p-value of $p < .05$ on cluster level are reported as significantly activated. This corresponds to a minimal cluster volume of approximately 1200 mm^3. The same statistical analysis was also executed for the regressor that represented the reading of the last six words of the context sentence.

Table 3 presents the 13 contrasts that were calculated in the whole brain analysis. The most important comparisons are set in bold font. These contrasts are directly connected to the experimental hypotheses (contrasts number C1, C5, C6, and C7). For all comparisons between "real word" sentence conditions (this excludes comparisons to pseudowords) contrasts were calculated "in both directions", such that each condition served as minuend and subtrahend respectively. The contrasts specified in Table 3 reflect the direction in which the indicated effects were expected.

The first four comparisons referred to the regressor which represented the reading of the context sentences. Contrast C1 (Real words > Pseudowords) was used to identify the extended language network by displaying cortical activity due to the average of explicit, paraphrase, inference, and unrelated condition minus activity due to reading pseudowords. For C2 (Paraprases > Explicit) and C4 (Unrelated > Inference) no differences were expected, because reading and understanding these sentences should not differ in processing requirements. These comparisons were only included to validate this assumption and the effectiveness of the sentence material. Significant results in C3 (Inference > Paraphrase) would indicate that the "open-ended" inference sentences triggered elaborative processes which are not necessary in paraphrase sentences as these mention the outcome of the specific event explicitly.

Table 3: Overview of contrasts analyzed in Experiment 1. The comparisons C1-C4 refer to the reading regressor, C5-C13 refer to the test regressor. For all pairwise comparisons between explicit, paraphrase, inference, and unrelated conditions both directions, "greater than" and "smaller than", were tested.

No	Phase of experiment	Conditions	Description
C1	Reading	**Real words > Pseudowords**	Extended language network
C2	Reading	Paraphrase > Explicit	Validation (no differences predicted)
C3	Reading	Inference > Paraphrase	Exploration of online generation of inferences
C4	Reading	Unrelated > Inference	Validation (no differences predicted)
C5	Test	**Paraphrase > Explicit**	Evaluation of lexical meaning (propositional level processing)
C6	Test	**Inference > Paraphrase**	Verification based on situation level processing (inferencing)
C7	Test	**Unrelated > Inference**	Exploration (different responses)
C8	Test	Inference > Explicit	Exploration (different baselines)
C9	Test	Unrelated > Explicit	"
C10	Test	Explicit > Pseudowords	"
C11	Test	Paraphrase > Pseudowords	"
C12	Test	Inference > Pseudowords	"
C13	Test	Unrelated > Pseudowords	"

The contrasts C5-C13 refer to the verification test. C5 (Paraphrase > Explicit) and C6 (Inference > Paraphrase) stood in the main focus of this study and have been described in detail above. C7 (Unrelated > Inference) should reveal processes that are related to the rejection of the unrelated statement based on situation level processing.

C8-13 were included for exploratory purposes to compare each of the sentence conditions to two kinds of baseline conditions. These were, on the one hand, comparisons against the explicit condition as baseline C8 (Inference > Explicit) and C9 (Unrelated > Explicit), and comparisons against the pseudoword condition as baseline (C9-C13). The latter contrasts of real word conditions and the pseudoword condition were mainly included to provide a baseline that was identical in Experiment 1 and Experiment 2.

For all visualizations of brain activations, statistical activation maps were either superimposed on inflated brain surfaces (PALS-B12) using the Caret software from http://brainvis.wustl.edu/caret (Van Essen et al., 2001; Van Essen, 2002), or visualizations were overlays of activations with sections of an averaged anatomical image of all participants from Experiment 1. Overlays with sections were used when activations occured deeper below the brain surface, so that the projection onto the brain surface template were not informative. Overlays were created with SPM2 and xjview (http://people.hnl.bcm.tmc.edu/cuixu/xjView/).

B.1.5.2. Regions of Interest

In addition to the relatively conservative whole brain analysis, it was also investigated whether the sentence manipulations modulated activity in functionally defined regions of interest. This analysis has more statistical power to reveal regional effects. Regions of interest (ROI) were functionally defined using the group comparison of reading real words (explicit, paraphrase, inference, and unrelated condition) versus reading pseudowords (see Table 6, Figure 7). This contrast was chosen as it was expected to identify major components of the extended language network. Within the five clusters that showed significantly differential activation, we chose all local maxima that surpassed a threshold of $z > 4.0$ and were at least 18 mm apart from each other. A sphere of 10 mm radius was drawn around each local maximum and intersected with the original cluster. This resulted in ten ROIs (see Table 4, Figure 6). Statistical analysis within the ROIs was conducted using the MARSBAR toolbox for SPM (Brett, Anton, Valabregue, & Poline, 2002) which averages over all voxels within a ROI. The modelling was exactly the same as the one used for the whole brain analysis, the only difference being the correction for multiple comparison. For the ROI analysis a Bonferroni-correction for the ten ROIs was applied.

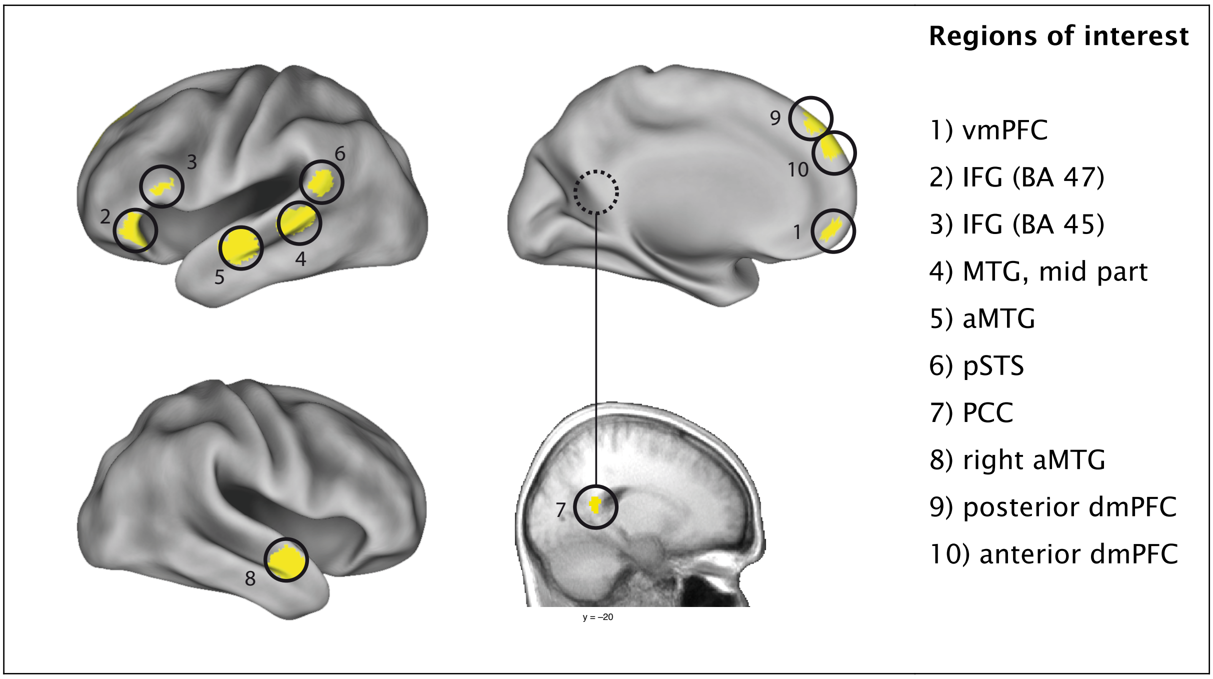


Figure 6: Regions of interest in Experiment 1. ROIs were functionally defined on the basis of the results from the "Reading words > Reading pseudowords" comparison. The left part of the figure depicts the lateral ROIs of the left (top) and right (bottom) hemispheres. On the right side ROIs at the anterior mid-line are projected onto the medial surface of the left hemisphere (top) and the posterior ROI at the PCC is shown as overlay on a sagital section of the participants´ averaged anatomical image (bottom).

Only a subset of the contrasts listed in Table 3 was considered in the ROI analysis. These contrasts included C5 (Paraphrase > Explicit), C6 (Inference > Paraphrase), and C7 (Unrelated > Inference) as well as the additional comparisons of inference and unrelated condition against the explicit condition as a common baseline (C8 and C9).

To illustrate the time courses of activation in the different sentence conditions finite impulse response (FIR) models were used (Ollinger, Corbetta, & Shulman, 2001; Ollinger, Shulman, & Corbetta, 2001). FIR modelling allows the deconvolution of the BOLD signal without assuming a particular response shape, which is known to vary across brain regions and participants (Handwerker, Ollinger, & D'Esposito, 2004).

Time courses of the BOLD activation, starting with the presentation of the headline and lasting 28 seconds, were extracted for each participant in each ROI for the four sentence conditions. Each time course was normalised to percent change and then averaged over participants separately for each ROI and condition. All calculations were done with MARSBAR (Brett et al., 2002) and Matlab (The Mathworks, Inc., Natick, MA, USA).

Table 4: Regions of interest in Experiment 1. ROIs were functionally defined on the basis of the results from the “Reading words > Reading pseudowords” comparison. Spheres of 10 mm radius around local maxima were intersected with the original clusters. Refer to the Methods section for details.

ROI	Description	L/R	Brodmann Area	MNI-Coordinates	Size in mm^3
1	Ventromedial prefrontal cortex	L/R	10	2, 56, -14	1432
2	Inferior frontal gyrus	L	47	-44, 32, -12	2592
3	Inferior frontal gyrus	L	45	-50, 18, 14	368
4	Middle temporal gyrus, mid part	L	21, 22	-54, -38, 2	2760
5	Anterior middle temporal gyrus	L	21	-58, -12, -14	3160
6	Posterior superior temporal gyrus	L	22, 40	-60, -50, 18	2144
7	Posterior cingulate cortex	L	30	-20, -52, 14	696
8	Anterior middle temporal gyrus	R	21	56, 0, -18	3120
9	Dorsomedial prefrontal cortex	L	9, 8	-8, 44, 46	2256
10	Dorsomedial prefrontal cortex	L	10, 9	-10, 56, 32	2936

B.2. Results

B.2.1. Behavioral Results

For all statistical tests α = .05 was assumed, and in pairwise comparisons α was adjusted according to the Bonferroni-Holm procedure. An ANOVA revealed that the participants´ response times in correct trials differed significantly across sentence conditions (explicit, paraphrase, inference, unrelated, and pseudowords), $F(4,48) = 20.5$, $p < .001$. The effect of the sentence conditions was also significant for the response frequencies, $F(4,48) = 6.3$, $p < .001$. The response latencies increased descriptively from the pseudoword condition to explicit, paraphrase, inference, and unrelated condition. The performance in inference and unrelated conditions was slightly lower than in the other conditions (see Table 5).

Planned pairwise comparisons were carried out for the following conditions: (1) pseudowords versus explicit, (2) explicit versus paraphrases, (3) paraphrases versus inferences, and (4) inferences versus unrelated. The response time difference between the pseudoword and explicit condition, $t(12) = 3.08$, $p = .01$ reached significance. All other pairwise comparisons were not significant on the corrected α-levels: explicit versus paraphrases, $t(12) = 1.19$, $p = .26$; paraphrases versus inferences, $t(12) = 2.25$, $p = .04$; inferences versus unrelated, $t(12) = 2.56$, $p = .03$.

Table 5: Average response times and accuracies from Experiment 1. In the explicit, paraphrase, and inference condition only "yes"-responses were included, whereas in the unrelated condition "no"-responses were considered. For the pseudoword condition all correct responses were analyzed.

	Response times in ms (SE)	Rel. frequency (SE)
Pseudowords (correct)	828 (37)	1 (.00)
Explicit ("Yes")	961 (54)	.99 (.01)
Paraphrases ("Yes")	999 (57)	.98 (.01)
Inference ("Yes")	1085 (61)	.89 (.04)
Unrelated ("No")	1207 (65)	.90 (.02)

B.2.2. fMRI Results: Whole Brain Analysis

The results of the whole brain fMRI analysis are depicted in three sections. First, the main contrasts for the reading and test phase of the experiment are reported (corresponding to C1-C7 in Table 3). These results are then supplemented with comparisons in which activity due to explicit statements is also subtracted from inference and unrelated statements (C8 and C9 in Table 3), and finally the sentence verification statements are compared to the pseudoword task (C10-C13 in Table 3).

B.2.2.1. Reading Regressor

Reading Pseudowords > Reading Real Words. A widespread network of primarily left hemisphere areas was more activated during the reading of the context sentences (explicit, paraphrase, inference, and unrelated conditions) than during reading the pseudoword sequences (Table 6, Figure 7). Activation was found in the left middle and superior temporal gyri, extending into the left inferior parietal lobe. Further clusters were located in the left inferior and middle frontal gyri, the medial parts of left superior frontal gyrus, and a region near the left posterior cingulate cortex and the precuneus. Moreover, there were active areas in the right middle, inferior, and superior temporal gyri.

Table 6: Specification of clusters from the "(C1) Reading words > Reading pseudowords" comparison of Experiment 1. A cluster-level threshold of $p < .1$ was applied. MNI-coordinates denote the voxel with highest z-value within the cluster.

Contrast / Region	Side	Brodmann area (BA)	Size (mm^3)	p_{corr}	z_{max}	MNI-coordinates (x, y, z)
1. MTG, IFG, STG	L	21, 47, 22	2,776	< .01	5.74	-58, -12, -14
2. Dorso-medial PFC	L/R	9, 10, 8	8544	< .01	5.11	-10, 56, 32
3. MTG, ITG, STG	R	21	4760	< .01	5.47	56, 0, -18
4. Ventro-medial PFC	L/R	11	3120	< .01	4.53	2, 56, -14
5. PCC	L	29	1216	.03	4.32	-20, -52, 14

Reading: Comparisons of Sentence Conditions (C2-C4). None of the predefined contrasts showed significant results with respect to the reading phase of the experiment. Ahead of the discussion it can be said that this result is not surprising, given that only the comparison between inferences and paraphrases was associated with a specific (weak) hypothesis about potential elaborative inference processes.

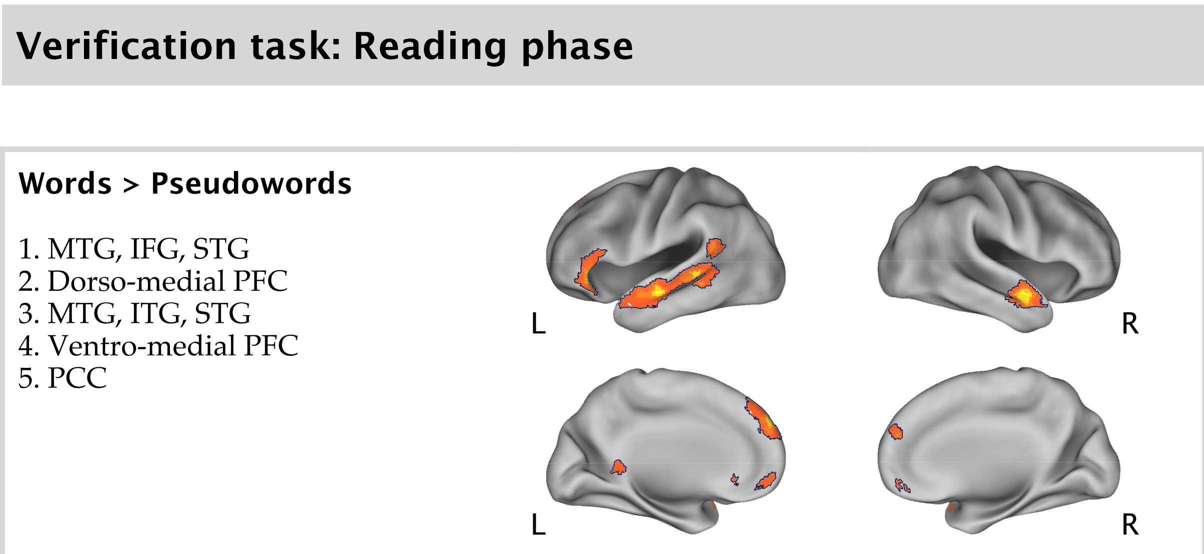


Figure 7: Significant clusters from the "(C1) Reading words > Reading pseudowords" comparison of Experiment 1 visualized on lateral and medial surfaces of inflated brain templates.

B.2.2.2. Verification Regressor

Main contrasts. Concerning the verification test and the primary interest of this study the following main results were found. Although not significant according to the conventional criterion, there was a tendency ($p = .06$) for an area in the right posterior cingulate gyrus and precuneus that was more activated in the paraphrase than in explicit condition (C5). In the contrast of inference versus paraphrase condition (C6), one cluster in the bilateral superior and medial frontal gyri (dmPFC), corresponding to Brodmann areas (BA) 8 and 9, was significantly activated (Table 7, Figure 8). No significant differences were found between unrelated and inference condition (C7).

Table 7: Results of the main contrasts from the whole brain analysis of Experiment 1. Shown are specifications of clusters with cluster-level p < .1. MNI-coordinates denote the voxel with highest z-value within the cluster. Reverse comparisons concerning the verification test regressor (explicit > paraphrase, paraphrase > inference, inference > unrelated) did not result in any significant clusters.

Contrast / Region	Side	Brodmann area (BA)	Size (mm^3)	p_{corr}	z_{max}	MNI-coordinates (x, y, z)
(C5) Paraphrase > Explicit (n.s.)						
PCC, cuneus	R	30	904	.06	3.92	10, -60, 8
(C6) Inference > Paraphrase						
Dorsomedial PFC	L/R	8/9	5472	< .01	4.91	-2, 46, 46
(C7) Unrelated > Inference						
No significant activation	-	-	-	-	-	-

Explicit condition as baseline. Comparing inferences against explicit statements (C8) resulted in two significant clusters (Table 8, Figure 9). A large left lateral cluster comprised the IFG and overlapped with the aTL. The second cluster was located in a bilateral frontal mid-line region, extending along the dmPFC, ACC, and SMA.

C9 (Unrelated > Explicit) yielded wide spread activity in both hemispheres. Bilateral activity was found in the dmPFC, vlPFC, and dorso-lateral regions around the precentral gyri. In the left hemisphere large portions of the IFG were activated, overlapping with the aTL. Further clusters were located in the left MFG, near the precentral gyrus, and in the left pTL.

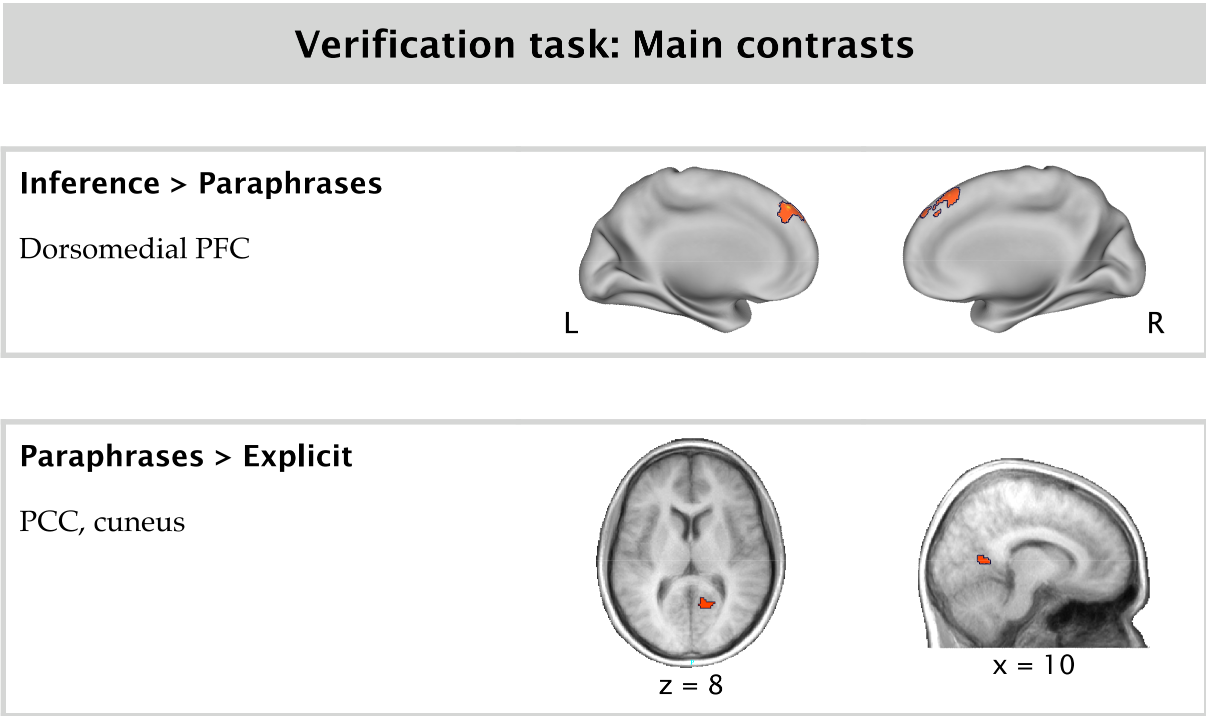
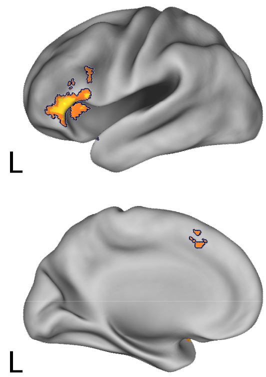


Figure 8: Results from the main contrasts of Experiment 1. Verifying inference statements resulted in greater activation than verifying paraphrase statements in the dmPFC. The processing of paraphrase statements yielded more activation than explicit statements in the PCC.

Verification task: Explicit condition as common baseline

Inference > Explicit

1. IFG, temporal pole, insula
2. SMA, ACC, dmPFC



Unrelated > Explicit

1. dmPFC
2. MTG, STG
3. IFG , aTL
4. precentral gyrus, SMA
5. pCC
6. pMFG
7. IFG

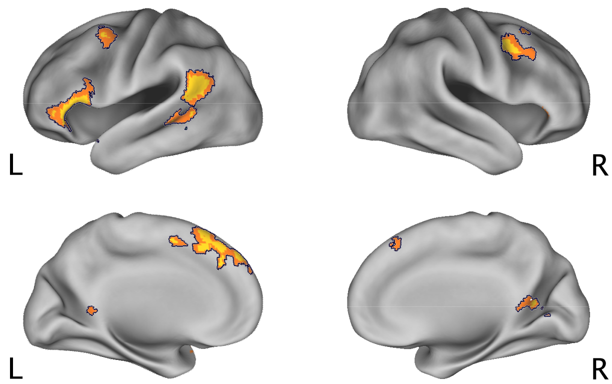


Figure 9: Comparisons of the unrelated and inference conditions with the explicit condition as common baseline (C8, C9).

Table 8: Whole brain analysis: inference and unrelated conditions versus explicit condition as a common baseline. Shown are specifications of clusters with cluster-level $p < .1$. MNI-coordinates denote the voxel with highest z-value within the cluster.

Contrast / Region	Side	Brodmann area (BA)	Size (mm^3)	p_{corr}	z_{max}	MNI-coordinates (x, y, z)
(C8) Inference > Explicit						
1. IFG, temporal pole, insula	L	47	1071	< .001	4.76	-40, 14, -18
2. SMA, aCC, dmPFC	L/R	8, 32	170	.010	3.64	-4, 18, 48
(C9) Unrelated > Explicit						
1. dmPFC	L/R	9, 8	1092	< .001	4.32	-10, 26, 60
2. MTG, STG	L	22, 39, 40	967	< .001	4.53	-52, -54, 22
3. IFG (pars orbitalis), aTL	L	45, 47, 38	933	< .001	4.26	-54, 24, 4
4. precentral gyrus, SMA	R	9, 8	397	< .001	4.50	44, 4, 44
5. pCC	L/R	30	272	< .001	4.46	18, -56, 14
6. pMFG	L	6	217	.002	4.39	-38, 0, 44
7. IFG	R	47	167	.008	3.91	52, 26, -6

Pseudoword condition as baseline. Using the pseudoword condition as a baseline the following results were found (Table 9, Figure 10). First of all, there were no areas in which explicit statements lead to significantly more activity than pseudoword statements (C10). One cluster in the left anterior temporal lobe (MTG/STG) showed more activation in the paraphrase condition than in the pseudoword condition (C11). More activity in the inference condition than in the pseudoword condition (C12) was associated with six regions. In left lateral cortex areas, there was one cluster in the IFG that overlapped with activation in the aTL (1), and a second cluster in the posterior temporo-parietal area (2). Three clusters were found in midline areas: One cluster in the left dmPFC, covering parts of BA 9/10 (3), a further cluster was located in the bilateral SMA (4), and one in the posterior cingulate gyrus, bilaterally (5). An additional activated area in the right hemisphere was located in the anterior temporal lobe (6). Unrelated statements produced more activity than pseudoword statements (C13) in the dmPFC (1), in the left IFG (2), along the entire left superior temporal sulcus (3), in the left MFG (4), in an area that partially overlapped with the posterior cingulate gyrus (5), and finally in cluster in the right anterior MTG (6).

Table 9: Specification of clusters from the comparisons with the pseudoword condition as a common baseline in the whole brain analysis of the verification regressor (Experiment 1).

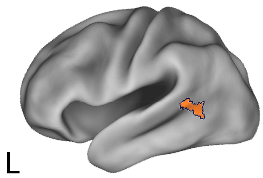
Contrast / Region	Side	Brodmann area (BA)	Size (mm^3)	p_{corr}	z_{max}	MNI-coordinates (x, y, z)
(C10) Explicit > Pseudowords						
no significant clusters	-	-	-	-	-	-
(C11) Paraphrase > Pseudowords						
pMTG/STG	L	22, 39	1872	<.01	3.85	−50 −48 8
(C12) Inference > Pseudowords						
1. IFG, MFG, aTL	L	47, 45, 38	11064	<.01	4.98	−52 24 −8
2. pMTG/STG, supramarginal gyrus	L	21, 22, 40	8912	<.01	4.92	−50 −46 8
3. dmPFC	L	9, 10	2536	<.01	4.49	−12 60 34
4. SMA	L/R	6, 8	2176	<.01	4.07	−6 38 56
5. Corpus callosum, pCC	L/R	29, 30	1040	.02	4.03	−12 −42 8
6. aMTG	R	21	968	.03	3.82	48 6 −24
(C13) Unrelated > Pseudowords						
1. dmPFC	L/R	8, 9, 6	8184	<.01	5.11	−8 58 38
2. IFG/MFG	L	47, 45, 44	9672	<.01	5.00	−44 32 −10
3. MTG/STG	L	21, 22, 38	14048	<.01	4.97	−36 16 −24
4. MFG	L	6	2664	<.01	4.68	−36 8 54
5. Limbic lobe, pCC	L/R	29, 30	3184	<.01	4.67	−12 −40 8
6. aMTG	R	21	1440	<.01	4.12	56 −4 −20

Note. Shown are specifications of clusters with cluster-level $p < .05$. MNI-coordinates denote the voxel with highest z-value within the cluster.

Verification regressor: Pseudoword condition as baseline

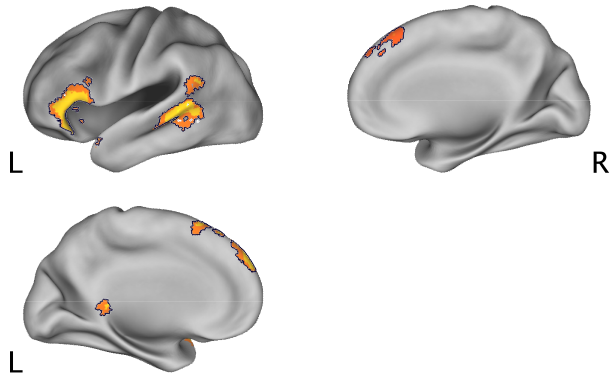
Paraphrases > Pseudowords

pMTG/STG



Inferences > Pseudowords

1. IFG, MFG, aTL
2. pMTG/STG, SMG
3. dmPFC
4. SMA
5. Corpus callosum, PCC
6. aMTG



Unrelated > Pseudowords

1. dmPFC
2. IFG/MFG
3. MTG/STG
4. MFG
5. Limbic lobe, pCC
6. aMTG

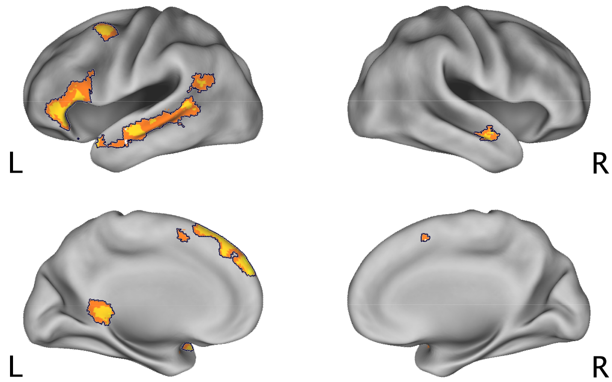


Figure 10: Comparisons with the pseudoword condition as a common baseline in the whole brain analysis of the verification regressor (Experiment 1). No significant clusters were found for the “Explicit > Pseudowords” contrast.

B.2.3. fMRI Results: Regions of Interest Analysis

The results of the regions of interest analysis are presented in Table 10, Figures 11, 12, and 13. Again, there was no area in which paraphrases elicited significantly more activation than explicit statements.

Table 10: Results of the ROI analysis concerning the test regressor. The GLM-based statistical analysis revealed significant results only for the reported three comparisons. Average activations were analyzed using *t*-statistics. *p*-values were corrected for multiple comparisons using the Bonferroni-procedure.

ROI	Area	Inference > Paraphrase		Inference > Explicit		Unrelated > Explicit	
		t	p_{corr}	t	p_{corr}	t	p_{corr}
1	L/R-ventromedial PFC	-	-	-	-	2.86	.07
2	L-ventrolateral PFC	2.95	.06	4.88	<.01	5.78	<.01
3	L-IFG	-	-	4.12	<.01	4.53	<.01
4	L-MTG, mid part	-	-	3.57	.02	4.58	<.01
5	L-anterior MTG	-	-	-	-	3.75	.01
6	L-posterior STG	-	-	2.97	.06	7.33	<.01
7	L-posterior CC	-	-	-	-	-	-
8	R-anterior MTG	-	-	3.15	.04	4.51	<.01
9	L-dorsomedial PFC	4.49	<.01	3.33	.03	4.58	<.01
10	L-dorsomedial PFC	3.01	.05	3.04	.05	4.25	<.01

Likewise, the unrelated condition was not dissociated from the inference condition. The inference > paraphrase contrast was significant in the posterior dmPFC (ROI 9) and close to significance in the more anterior dmPFC (ROI 10) as well as in the left anterior IFG (ROI 2). Descriptively, the activation time courses indicate that unrelated and inference condition on the one hand, and explicit and paraphrase condition on the other, elicited similar activation patterns in most regions of interest. Except for the posterior cingulate area (ROI 7), unrelated and inference trials seemed to be associated with more activation than paraphrase and explicit trials. This observation was confirmed by contrasting unrelated and inference condition with the explicit condition as common baseline. Unrelated trials produced significantly more activation than explicit trials in all regions of interest except for ROI 7. The inference condition was associated with more activation than the explicit condition in the left IFG (ROI 2 & 3), in the dmPFC (ROI 9), and in the left and right MTG (ROI 4 & 8).

Estimated activation time courses of the different conditions were included to illustrate and supplement the inferential statistics presented above. Time course data in general reveals information about the shape of the hemodynamic response to different stimuli, and can thus inform about temporal aspects of the response and also about the polarity of the response (activation vs. deactivation). Furthermore, the analysis can serve as a validation of the regressor-based analysis.

With some exceptions, the following generalizations can be made across regions of interest: It appeared that time courses of the unrelated and inference condition followed similar patterns on the one hand, and time courses of paraphrase and explicit condition on the other. Overall, the shapes of the responses within a ROI were comparable across conditions, differing mainly in response amplitudes. The unrelated and the inference condition tended to cause highest BOLD responses, followed in ranking by the paraphrase, and explicit conditions. The difference between the amplitudes of inference condition and paraphrase condition was greater than differences between explicit and paraphrase condition, or between inference and unrelated condition, respectively. This was in particular the case at the peak amplitudes. Within a ROI, the differences between the latencies of reaching peak amplitudes in the different conditions, were usually not greater than 2 seconds. In most regions the response functions tended to be bimodal, with a global maximum 14 to 18 seconds after trial onset, and a second local maximum, approximately 4 seconds after trial onset. These early local maxima were significantly lower than the global maxima, and in some ROIs there was only a local increase of the response function gradients at this time point (e.g., ROI 6).

Left lateral regions of interest. Time courses in the left ventrolateral PFC (ROI 2) follow the general pattern expressed above. With exception of the paraphrase condition, the time courses of all conditions show a local maximum at 4 seconds. In the explicit and paraphrase conditions, time courses reach their global maximum after 16 seconds, whereas in the unrelated and inference condition peaks occurred at 18 seconds. From 16 seconds on, the unrelated and inference condition lead to higher amplitudes than the explicit and paraphrase condition. In the second left frontal region of interest (ROI 3, BA 45), response shapes were unimodal, with global maxima between 16 and 18 seconds. Again, descriptively the time courses of the inference and unrelated conditions deviated from those of explicit and paraphrase condition (most noticeable between 18 and 22 seconds.).

In the two left posterior temporal lobe regions (ROI 4 and ROI 6) similar response pattern were observed. There was a noticeable increase of the response function gradients between 2 and 4 seconds, although no clear local maxima occured. Global maxima emerged around 16 seconds. The differences between the unrelated and inference conditions on the one hand, and explicit and paraphrase conditions on the other hand, were not as pronounced as in the frontal regions. In the anterior temporal lobe region (ROI 5), these differences were marginally recognizable, and the response functions were more shallow with lower maximum amplitudes.

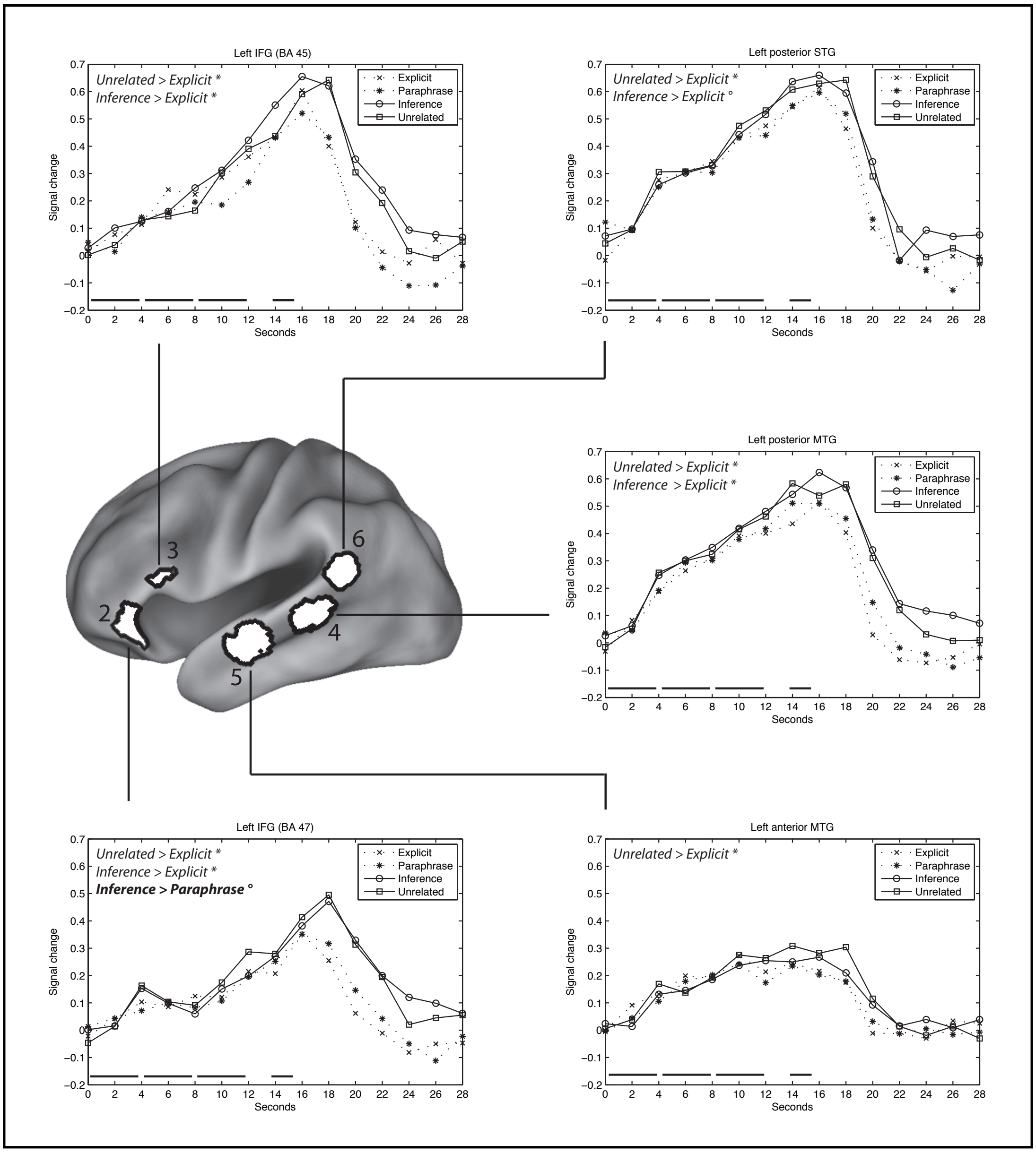


Figure 11:
Results of ROI analysis in left lateral areas. ROIs were projected on the cortical surface of an inflated brain template. Time course diagrams show the estimated signal change in each ROI for the four sentence conditions from the onset of the headline to the end of the trial. Horizontal lines just above the x-axis of each diagram approximately indicate the timing details of the GLM regressors for sentence reading and verification test. In the upper left corner of each diagram significant results with respect to the GLM verification regressor contrasts are indicated (* $p < .05$; ° $p < .1$). See Table 4 for specifications of the ROIs and Table 10 for details of the statistical analysis.

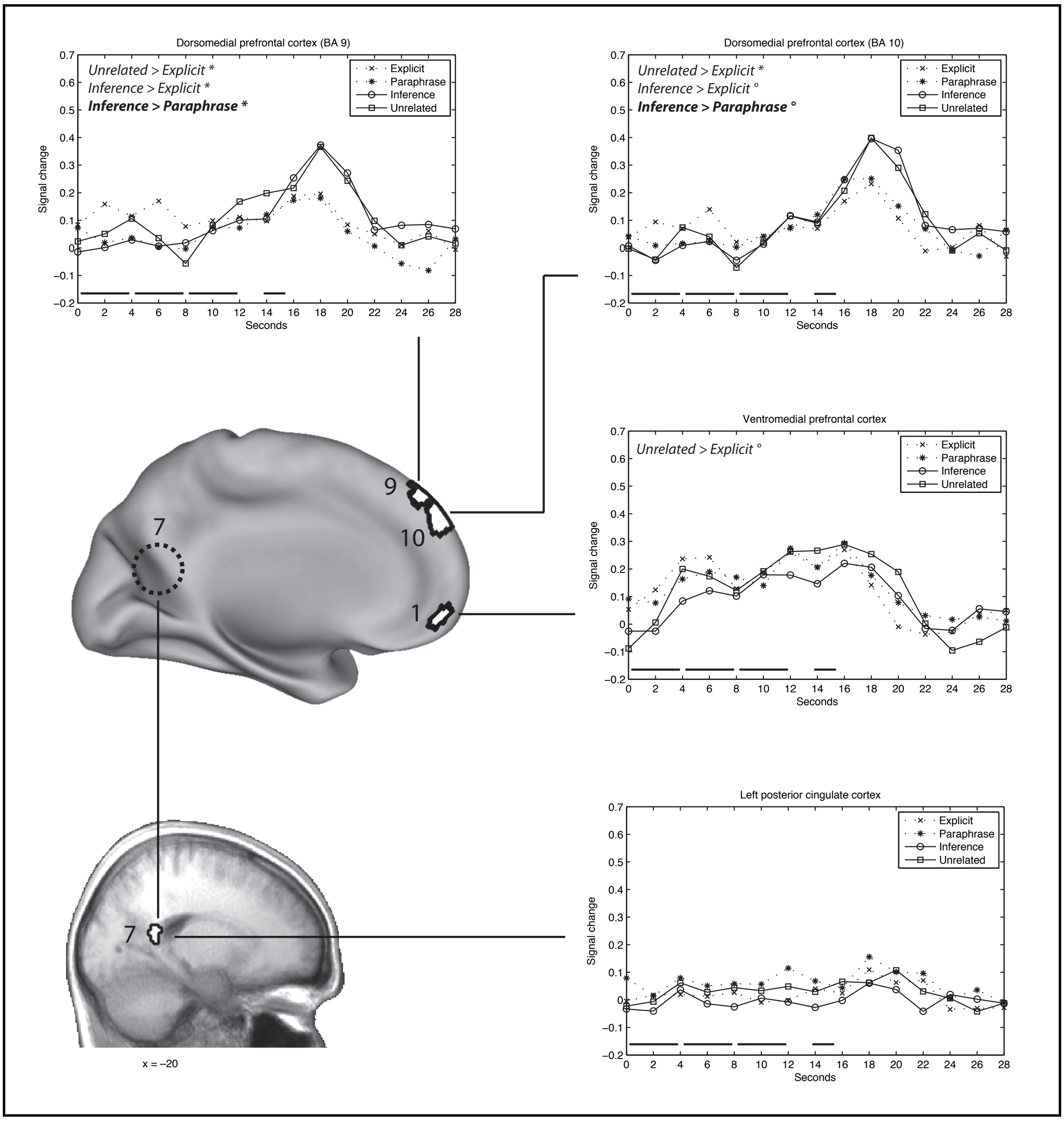


Figure 12:
Results of the ROI analysis in left medial areas. See caption of Figure 11 for more information.

Midline regions of interest. In comparison to the time courses in the left lateral regions (except for ROI 5 in the aTL), the time courses in the two adjacent dmPFC areas (ROI 9 and ROI 10) rised slower, and developed lower global maxima (after 18 seconds). Between 18 seconds and 20 seconds, the difference of the inference/unrelated conditions and explicit/paraphrase conditions was most pronounced. In the ventromedial region (ROI 1), the response shapes were broader, and global maxima were lower. Time courses in the posterior cingulate area (ROI 7) increased only marginally above zero throughout the whole trial length.

Right anterior temporal lobe. In ROI 8, the time courses roughly followed the general pattern, although on an overall low activation level. There was a tendency for an increase of activation after 4 seconds, and global maxima emerged between 14 and 16 seconds.

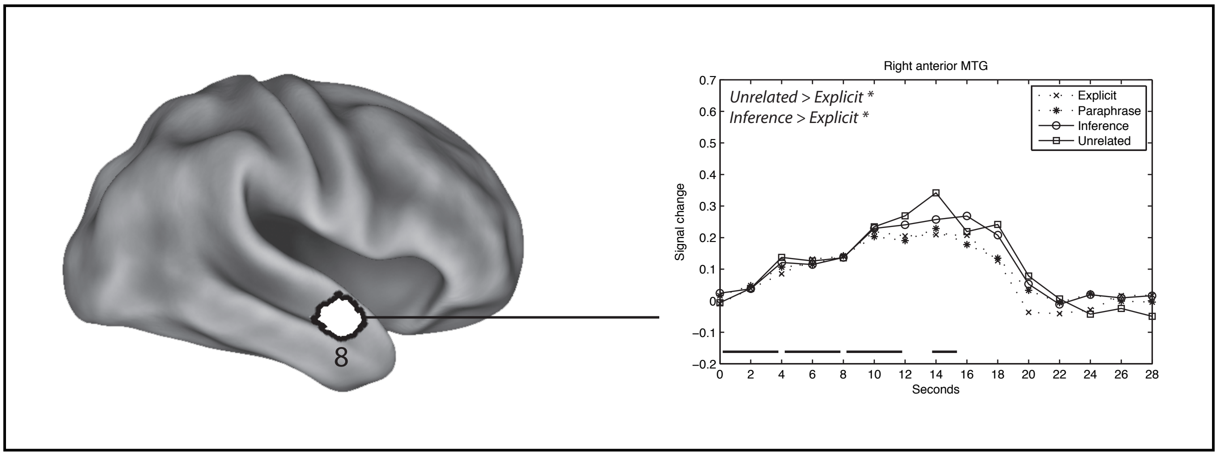


Figure 13: Results of ROI analysis in the right temporal lobe. See caption of Figure 11 for more information.

B.3. Discussion

In this experiment, participants verified short statements after reading a context sentence. In the inference condition, the verification statement could be related to the context statement by an inference process. Moreover, the design of the sentence material was guided by the theoretical notion that language is represented on different levels (McDaniel et al., 2001; Schmalhofer et al., 2002). Comparisons between the conditions therefore reflected the processing of the verification statements at the three different representational levels to varying degrees.

Many regions contributed to the language comprehension processes due to the reading of real sentences as opposed to the reading of pseudoword sequences (Figure 7). The identified network comprised areas which are frequently found in studies on language processing (Ferstl et al., 2007): portions of the left and right temporal lobes, the left IFG, as well as the mid-line areas dmPFC, vmPFC, and PCC.

In the paraphrase and explicit conditions the verification statements and context sentences supposedly shared the same situation level representation as well as the propositional representation, and they differed only in the surface representation. As the literal wording of the test statement was not decisive for the verification task, the average proportion of yes-responses for the explicit and the paraphrase condition did not differ significantly. While in the whole brain fMRI analysis one cluster in the right posterior cingulate cortex almost reached significance in the corresponding contrast of paraphrase and explicit condition, no significant differences were found in any of the regions of interest for this contrast. Activation in the right posterior cingulate and precuneus area (ROI 7) has been found in some studies on language comprehension and is assumed to reflect memory retrieval processes (Bottini et al., 1994; Ferstl & von Cramon, 2002).

Inference and paraphrase statements differed with respect to the surface level and the propositional level from the representations of the context sentences. As the verification tasks required an evaluation of the meaning of the test statement in relation to the context sentence, the lack of a corresponding propositional representation in the inference condition required an elaboration of the situation model. In the whole brain fMRI analysis it was found that the dmPCF was the only area where significantly more activity was found in the inference trials than in the paraphrase trials. The regions of interest analysis confirmed this result (ROI 9 & 10) and additionally pointed to a stronger involvement of the left inferior frontal gyrus (BA 47, ROI 2) in inference trials.

The representations of unrelated statements did not match the representations of the context sentences on neither level. Due to the verification instruction the participants had to evaluate whether the unrelated statement could be integrated into the situation model that was built according to the context sentence, and if not, the unrelated statement had to be rejected. The behavioral results showed no significant differences between these conditions. The interpretation of these results is difficult though, because unrelated statements required a no-response while only yes-responses were considered in the inference condition. Interestingly, the lack of significant activation in the comparison of unrelated and inference condition in the fMRI analyses seems to suggest that there was considerable overlap between the areas that were involved in the inference generation process on the one hand and in the evaluation of implausible statements on the other hand. This was also illustrated by the activation timecourse diagrams, that showed closely corresponding curves for inference and unrelated trials in most regions of interest.

The following discussion first proposes an interpretation of the results with respect to the brain networks which supported comprehension processes in the current task. Secondly, the results are compared to other imaging studies on inferencing in more detail.

B.3.1. Inferencing and the Prefrontal Cortex

According to McDaniel et al. (2001) and Schmalhofer et al. (2002) inference and paraphrase statements shared the same situation level representations with the corresponding context sentence but differed with respect to the propositional and surface level representations. Thus, more activity in the inference trials than in paraphrase trials revealed comparison processes at the situational level, which are thought to rely on the episodic memory system (Kintsch, 1988; Magliano, Radvansky, & Copeland, 2007). The most prominent region which could be associated with these processes was the dmPFC. There was one large significant cluster found in the whole brain analysis (Figure 8). In the ROI analysis significant activations were located in the posterior dmPFC (BA 9, ROI 9) and tendencies towards significance in the anterior dmPFC (BA 10, ROI 10) as well as in the left anterior inferior prefrontal cortex (ROI 2).

The finding that regions in the dmPFC seem to play a central role in inference processing is in close agreement with different lines of research. In the language domain, Ferstl and von Cramon (2001) have demonstrated that the dmPFC was activated when their participants evaluated the coherence of pragmatically related sentences. Mazoyer et al. (1993) for example also found activity in the medial aspects of the superior frontal gyrus comparing the processing of meaningful stories versus distorted stories. A recent meta-analysis also pointed to the importance of the dmPFC for establishing coherence (Ferstl et al., 2007).

Moreover, the dmPFC seems to be involved in a variety of other higher-level cognitive processes like theory of mind processing (e.g. Frith & Frith, 1999; Fletcher, Happe, Frith, & Baker, 1995; Gallagher & Frith, 2003), reasoning (Goel, Gold, Kapur, & Houle, 1997; Ruff, Knauff, Fangmeier, & Spreer, 2003), task-switching (Forstmann, Brass, Koch, & von Cramon, 2005), response inhibition (Li, Huang, Constable, & Sinha, 2006), and autobiographical and episodic memory (Cabeza & Nyberg, 2000a; Gilboa, 2004; Graham, Lee, Brett, & Patterson, 2003; Svoboda, McKinnon, & Levine, 2006). To embrace these diverse findings, Ferstl and von Cramon (2002) suggest that the dmPFC might have "domain-independent functionality related to volitional aspects of the initiation and maintenance of nonautomatic cognitive processes" (p. 1611).

Recently it has been proposed that this fronto-medial region is part of a network that generally subserves the reconstruction or simulation of past and future episodes based on prior experience (Addis, Wong, & Schacter, 2007; Buckner & Carroll, 2007; Schacter & Addis, 2007b). In this view the abilities to infer information and to predict events are core functions of the episodic memory system. In the field of language comprehension Zwaan (2004) has presented a related framework in which comprehension processes are also viewed as simulations at the situational level. On the basis of these ideas and the studies reviewed above, the activations which were found in the dmPFC seem to reflect the intentional evaluation of the plausibility of the verification statements based on the participants´ prior knowledge.

In addition to the significant activation in the dmPFC, the ROI analysis revealed some evidence for higher activation in the inference condition than in the paraphrase condition in the left anterior inferior PFC (ROI 2). Mostly based on sub-sentence-level studies it has been argued that the left inferior prefrontal cortex is critically involved in the selection of semantic representations (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) or the control of semantic retrieval (Wagner, Pare-Blagoev, Clark, & Poldrack, 2001). Jung-Beeman (2005) also highlighted the importance of this region for the selection of semantic concepts and thereby for the integration into a wider context. Similarly, Hagoort (2005) suggested that an important function of the left

IFG for language comprehension is ‘unification’—i.e. “integration of lexically retrieved information into a representation of multi-word utterances”. Kuperberg et al. (2003) for example showed that activity in these regions increases during the processing of pragmatically anomalous sentences relative to normal sentences. They suggest that a left temporofrontal language network serves in mapping and conceptually integrating words onto information in semantic memory.

In the present experiment increased activity in the anterior inferior PFC maybe also reflected the attempt to map and integrate inference and unrelated statements on a conceptual level. The absence of matching propositional and surface representations in these conditions probably required increased integration efforts. However, as the sentence material was not controlled for overall difficulty, the results remain somewhat inconclusive in this respect.

In addition to the activations in the dmPFC and left IFG, a third frontal area was identified in the comparison of real word reading and pseudoword reading, namely the vmPFC (BA 11). The ROI analysis furthermore revealed that only the comparison of unrelated and explicit statements was close to significance (p = .07) for this area. In the language domain, activation in the vmPFC has been associated with the processing of emotional aspects of situation models (Ferstl et al., 2005). Given that most of the sentences used describe situations which are likely to be accompanied by some emotions of the protagonists and may even elicit affective responses in the readers, this interpretation of the vmPFC´s function would be compatible with the results from the reading phase. However, the finding that unrelated statements yielded higher activation than explicit statements suggests that the vmPFC might instead (or in addition) contribute in other ways to language comprehension. For example, the vmPFC is also thought to play a role in the monitoring of memory contents and in *feeling of knowing* judgements (Schnyer, Nicholls, & Verfaellie, 2005). Schneyder et al. (2005) suggest that the vmPFC participates in making judgements about the fit of available evidence to a mental model of relevant context. This description clearly bears large similaritiy to the functions that have been ascribed to the dmPFC. In addition, in the meta-analysis of Ferstl et al. (2007) the vmPFC was found to be reliably associated to coherence building. However, in the present study, the difference between inference and paraphrase statements did not modulate activity in the vmPFC, but it did so in the dmPFC. Consequently, it seems that in the context of the verification task the dmPFC was more related to inference processing than the vmPFC.

B.3.2. Temporal and Parietal Areas

The critical contrast between inference and paraphrase condition did not yield significant results in temporal areas—neither in the whole brain analysis, nor in the regions of interest analysis. However, the regions of interest analysis provides evidence that these temporal areas did not respond completely undifferentiatedly to the test statements. In all temporal areas (ROI 4, 5, 6, 8)—including the right anterior MTG—the unrelated condition elicited more activity than the explicit condition. In the left middle portion of the MTG (ROI 4) the contrast between inference and explicit condition was also significant, and in the posterior left STG (ROI 6) this contrast was close to significance.

Most probably the modulations of temporal activity in the anterior and middle temporal lobes reflect more intense semantical processing of unrelated and inference statements, as compared to explicit statements, because the former are not part of the propositional representation of the context sentence. Of course, unrelated and inference statements also differ with respect to the surface representation from explicit statements. As the paraphrase condition did not elicit higher activation than the explicit condition in any of our regions of interest, it can be suspected that the modulations of activity in temporal areas were rather due to semantical processes. The involvement of areas in the anterior and middle parts along the superior temporal sulcus in semantic processing of sentences and discourse is suggested by numerous studies (Bottini et al., 1994; Ferstl & von Cramon, 2001; Ferstl & von Cramon, 2002; Ferstl et al., 2005; Kuperberg et al., 2003; Mazoyer et al., 1993; St George, Kutas, Martinez, & Sereno, 1999; Vandenberghe, Nobre, & Price, 2002; Virtue et al., 2006). Furthermore, like in the present study, activations in the anterior and middle parts of the temporal lobes often occur bilaterally during sentence comprehension (Bottini et al., 1994; Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Ferstl, 2007a). Ferstl et. al (2005) also found the right anterior temporal lobe to be activated more during the processing of inconsistent versus consistent information. They interpreted this finding as an indication for increased difficulty with the integration of the current information into the situation model.

Posterior left superior temporal areas, overlapping with regions in the inferior parietal lobe (supramarginal gyrus), have been associated with phonological processing in the context of verbal working memory (e.g., Gernsbacher & Kaschak, 2003; D'Arcy, Ryner, Richter, Service, & Connolly, 2004; Gitelman et al., 2005). The pattern of activity we found in this area (ROI 6) might also reflect differences in the requirements of verbal working memory in our conditions. Although speculative, it seems possible that the significant differences in the posterior temporal lobe between unrelated and explicit condition, as well as between inference and explicit condition might be caused by the participants´ attempts to rehearse the context sentences. Particularly as the presentation of the context sentences was word-by-word, rehearsal might have been more necessary if the test statements differed from the surface and propositional representation of the context sentences.

B.3.3. Comparison to Other Imaging Studies on Inferencing

As has been outlined in the introduction (section A.2.3), direct comparisons with prior imaging studies are necessarily selective because of differences in tasks, materials, and analysis methods. On a large scale, the present study used a modelling approach which is similar to that of Kuperberg et al. (2006) who also defined separate regressors for context sentences and critical test probe in an HRF-model, and who also additionally employed FIR-models. The pattern of results found in the present study is in part compatible with the results of Kuperberg et al. (2006). In their study the comparison of intermediately related sentences > highly related sentences was supposed to be strongly related to the drawing of causal inferences. Activations associated with this contrast were found in the IFG (BA 47) bilaterally, in left MFG, in IPL bilaterally, in left MTG, and in the medial PFC bilaterally. This overlaps with the dmPFC and (left) IFG activations found in the present study in the comparisons of inference > paraphrase and inference > explicit statements respectively. The fact that Kuperberg et al. (2006) additionally identified temporal and parietal areas might be attributed to differences in the task and their more heterogeneous sentence materials—which is subject to the confound of inference processing and general situation model building discussed in the introduction. In a similar manner the current study revealed a wide spread network of activations when comparisons were made between very different test probes in the unrelated > explicit comparison. Furthermore, the epoch the authors analysed with respect to the critical test word and the participants' responses was considerably longer (4.66 seconds) than the epoch used in the present study (1.8 seconds). Thus, the power to detect any effects might have been larger in the Kuperberg study but the price being less specificity because participants response times were on average only around 1 second.

The results of verification experiment are also compatible with several findings of Ferstl and her collegues. In Ferstl et al. (2001) the dmPFC and PCC were associated to inferencing on the basis of the comparison of coherent > incoherent sentence pairs (see section A.2.3 for an explanation why in this case this contrast supposedly revealed inference processes instead of the reversed contrast incoherent > coherent). The authors proposed that the PCC is involved during the integration of new information into the situation model. Again, the differences in the sentence materials can provide a reason why the PCC was not found activated in the inference > paraphrase contrast of the present experiment. In the latter the situation model representations of inference and paraphrase sentences overlapped to a large degree and differences occured only on the propositional and surface levels. Therefore, the inference verification probes did not constitute new information as the test sentences in the experiment by Ferstl et al. (2001).

Two further studies of Ferstl et al. (2002, 2005) provided evidence for an involvement of the dmPFC during inferencing. Furthermore, in Ferstl et al. (2005) the detection of inconsistencies was associated with the right anterior temporal lobe in the event-related analysis, and with the vlPFC bilaterally in the epoch-related analysis. The dmPFC showed significant activation in response to emotionally inconsistent stories but not to chronologically inconsistent stories. In the present study, activations in the right anterior temporal lobe and in the left vlPFC occurred in the comparisons inference > pseudowords and unrelated > pseudowords in the whole brain analysis, as well as in inference > explicit and unrelated > explicit in the ROI analysis. No evidence for an involvement of the right vlPFC, as shown by Ferstl et al. (2005), was found, which might be explained by the greater demands on semantic retrieval processes in the Ferstl study due to their longer, multi-sentence text materials.

The results of the two inferencing studies which advocated a predominant role of the right hemisphere during inferencing are are only partially compatible with the outcome of the present experiment. The bilateral dlPFC, proposed by Mason and Just (2004) for the generation of inferences, were not significantly activated in any contrasts of the present study. As Mason and Just (2004) did not analyze the dmPFC and aTL it is impossible to compare results concerning these regions. Virtue et al. (2006) proposed that the right STG is involved in the generation of inferences at a relatively early time point during discourse comprehension—when a verb requires an inference—, whereas the left STG is responsible for "later" inferences when a coherence break occurs. Both the left and right temporal activation have also been found in the present study, although the right cluster found by Virtue et al. (2006) was apparently located posteriorly from the aTL activation found in the present experiment and in the study of Ferstl et al. (2005). Modulation of activity in the left posterior STS region also occurred in several contrast of the whole brain analysis of the present study, i.e. in the unrelated > explicit, paraphrase > pseudowords, inferences > pseudowords, and unrelated > pseudowords comparisons. Additionally, the inferences > explicit and unrelated > explicit comparisons were significant in the ROI analysis. Importantly, the inference condition did not elicit greater activation than the paraphrase condition in this region. From the perspective of the present study it is somewhat puzzling why Virtue et al. (2006) found this posterior STG region to be the most prominent region which could be related to inference generation when coherence breaks occurred. Apparently the different experimental approaches differentially emphasize semantic activation processes in the case of the reading task of Virtue et al (2006) and evaluation processes in the case of the verification task.

B.3.4. Summary and Outlook

In this study, inference processing was studied with an event-related fMRI experiment based on theoretical assumptions derived from the text comprehension framework of Kintsch and van Dijk (Kintsch & van Dijk, 1978; Kintsch, 1998). The processing of context sentences and verification statements which contained the same situation model representation and differed in propositional and/or surface (explicit, paraphrase, and inference condition), with unrelated sentences that differed at all levels of representation were compared. Depending on the overlap of the different representations of context sentences and subsequently presented verification statements, differential activation patterns during the verification task could be observed. The fMRI data was firstly analyzed on the basis of pre-defined regressors in a whole brain analysis, and secondly an additional, more sensitive ROI analysis for a closer inspection of effects within regions was used. Furthermore, an independent, FIR-model based analysis was conducted to validate the results from the regressor-based analyses and to gain additional information about the time courses of activation.

In summary, it was shown that the dmPFC is critically involved in the generation of inferences from the situation model. This is compatible with results from several lines of research that highlight the importance of this area for the allocation of prior knowledge. By contrast, some other imaging studies on language comprehension did not find significant activation in the dmPFC, but rather point to a special role of right hemisphere areas for comprehension at the sentence or discourse level. In these studies participants typically read for comprehension, and they were not required to respond to an explicit task (e.g., Mason & Just, 2004). The results of the ROI analysis showed that the right anterior MTG actually did contribute to the additional processing that was required to verify inference and unrelated statements, in comparison to the explicit statements. However, the critical contrast between inference and paraphrase condition was not significant in this ROI. Thus, the present study provided some evidence for the involvement of the right anterior temporal lobe during inference processing, but it did not point to a pronounced function of this region in this task. Clearly, more research is needed to determine the exact conditions under which activities in the right hemisphere are found during language comprehension (see Ferstl, 2007a for a discussion of this topic).

C. Experiment 2: Recognition

The rationale of Experiment 2 followed the notion that the processing at the different representational levels can be enhanced selectively by manipulating task instructions (McDaniel et al., 2001; Zwaan & Singer, 2003). Participants were asked to perform a recognition task on the same sentence material that was used for Experiment 1. Griesel et al. (2003) also studied the effect of verification and recognition task instructions with similar materials. They estimated the strengths of the memory traces on different representation levels and found that the recognition instruction lead to a surface level representation that was twice as high as with the verification instruction ($d'_{\text{verification}}$ = .70 versus $d'_{\text{recognition}}$ = .26).

In contrast to Experiment 1, matching representations on the propositional and situational levels of context sentence and recognition statement can be viewed as a source of distraction or semantic interference. For the behavioral responses it follows that rejecting unrelated statements should be faster and more accurate than rejecting inference statements. The lowest performance levels can be expected for the paraphrase statements, which match the context sentence on situational and propositional level.

The interpretations of differential contrasts between the explicit, paraphrase, inference, and unrelated conditions in this study are less straightforward than in Experiment 1. In the recognition task participants were asked to respond *yes* only if the test statement was made up of words that had been presented already in the context sentence. In principle, this task can be accomplished without assessing the meaning of context sentence and recognition statements, because the visual or phonological word forms (surface level representations) carry sufficient information to reject paraphrase, inference, and unrelated statements. The model of sentence recognition and verification employed here (Kintsch et al., 1990) assumes that alle three levels of representation are used in the comparison process. Thus, for paraphrase statements the comparison process would evaluate that the wording of the statement is different from the wording in the context sentence, but also that the meaning of the words is the same. Similarly, for inference statements it can be assumed that the different wording is detected as well as the plausibility of the statement in the given context. Unrelated statements would be evaluated as differing in wording, and also as not fitting into the context. As paraphrase, inference, and unrelated statements are to be rejected, the rejection component can be expected to be cancelled out in pairwise differential contrasts of these conditions. Consequently, areas to be associated with greater activity during the processing of inference statements compared to paraphrase statements were assumed to indicate situation level processing like in Experiment 1. More activity due to unrelated statements than to inference statements would reflect the evaluation of the mismatch of representations on the situational level. As this comparison did not yield significant results in Experiment 1, even though different responses were required in the verification task, no significant differences between these conditions were expected in this study. Lastly, greater activity to paraphrase statements than to explicit statement should reflect rejection of the paraphrase statement based on an evaluation of the surface and propositional representations.

C.1. Methods

Most of the methodological details of Experiment 2 were identical to Experiment 1. Here, only the deviations are reported in detail, whereas the corresponding aspects are briefly summarized.

C.1.1. Design, Material, and Procedure

The same experimental stimuli as in Experiment 1 were used for this study with one exception. To achieve a better balance of expected yes- and no-response the context sentences and test statements of the filler trials were modified respectively. In Experiment 1 the filler trials were constructed to lead to increase the number of rejections (no-responses). While in Experiment 2 filler trials were used to increase the number of yes-responses, i.e. they were constructed to resemble explicit sentences. Overall, again 18 stimulus pairs consisting of a context sentence (or pseudoword sequence) and a test statement for each of the following conditions were used: explicit, paraphrase, inference, unrelated, filler, and pseudowords.

Except for the instructions the procedure was identical to Experiment 1. Participants were instructed to read each context sentence and the subsequent test probe, and to indicate with a button press as quickly and accurately as possible whether the words that made up the test statement were contained in the context sentence previously read. In the pseudoword condition the task was to decide if the test statement was identical to the last two pseudowords of the study sequence. There was a training session outside of the magnet and seven training trials inside the magnet, prior to the beginning of the actual experiment. For training, stimuli were used that were not included in the experiment.

After the presentation of the headline, context sentences and pseudoword sequences were again presented word-by-word. Two seconds after the last word, the test statement was presented until participants responded (within 4.3 seconds). After a variable inter-trial interval of 7-11 seconds the presentation of the next trial started. See Figure # for an overview of the timing details.

C.1.2. Participants

Fourteen participants volunteered in the fMRI study for course credit or payment. Six participants were female, and the average age was 25 years. All participants were right-handed, native speakers of German. They were healthy, had no history of neurological illness, and all had normal or corrected to normal vision (contact lenses). Informed written consent for participation in the fMRI study was obtained from all participants. Data from one participant was excluded from analysis because of extensive head movement that could not be corrected adequately (maximum of 18 mm y-translation).

C.1.3. fMRI Image Acquisition, Preprocessing, and Modelling

MR-images were again acquired at the MR-facility of the University of Oldenburg in the lab of Prof. Mark Greenlee with a 1.5 T Siemens Sonata whole body MRT. The same imaging sequences were used as in Experiment 1 (see p. 44 for details).

Table 11: Contrasts analyzed in Experiment 2. The comparisons C1-C4 refer to the reading regressor, C5-C14 refer to the test regressor. For all pairwise comparisons between explicit, paraphrase, inference, and unrelated conditions both directions, “greater than” and “smaller than”, were tested. The main contrasts of interest are set in bold face. C8 (Inferences > Explicit) and C9 (Unrelated > Explicit) which were included in Experiment 1 (Table 3) were not analyzed in Experiment 2. As unrelated and inference condition required rejection, the explicit condition was not appropriate as common base line condition.

No	Phase of experiment	Conditions	Description
C1′	Reading	Real words > Pseudowords	Extended language network
C2′	Reading	Paraphrase > Explicit	Validation (no differences predicted)
C3′	Reading	Inference > Paraphrase	"
C4′	Reading	Unrelated > Inference	"
C5′	Test	Paraphrase > Explicit	Exploration; rejection based on surface representation (same meanings, but different wording)
C6′	Test	Inference > Paraphrase	Situation level processing (statement is plausible, rejection component cancelled out)
C7′	Test	Unrelated > Inference	Situation level processing (statement is not plausible, rejection component cancelled out)
C10′	Test	Explicit > Pseudowords	Exploration (pseudoword baseline)
C11′	Test	Paraphrase > Pseudowords	"
C12′	Test	Inference > Pseudowords	"
C13′	Test	Unrelated > Pseudowords	"
C14	Test	Hits vs. Correct rejections	Recognition effects

Preprocessing and modelling of the fMRI data was in all aspects identical to Experiment 1. Separate regressors were used to account for the reading phase of the experiment and the test phase. Table 11 depicts an overview the contrasts that were analyzed. The stated direction of a comparison (e.g., Inference > Paraphrase) corresponds to the contrast of main interest. All “reversed” comparisons (e.g., Paraphrase > Inference) were also considered in the analysis.

C.1.4. Regions of Interest

Regions of interest were functionally defined with the same criteria used in Experiment 1. The group comparison of reading real words (explicit, paraphrase, inference, and unrelated condition) as opposed to pseudowords (C1′ in Table 11) was used to identify the brain areas that were active during reading comprehension. Spheres of 10 mm radius were built around local maxima and intersected with the original clusters. This resulted in total 17 regions of interest, specified in Table 12 and illustrated in Figure 14.

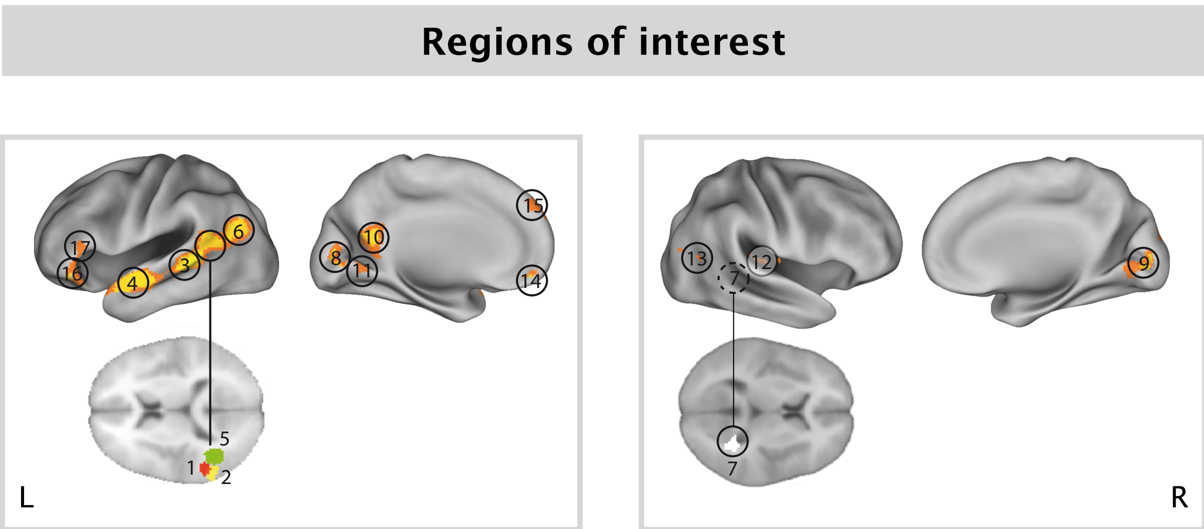


Figure 14: Regions of interest in Experiment 2. A total of 17 ROIs were functionally defined on the basis of the results from the “Reading words > Reading pseudowords” comparison. There was a minor overlap of thirteen voxels in ROI 1, 2, and 5 (of 1005 voxels in total). See Table 12 for more information on the regions.

Within the 17 ROI the same statistical GLM-based analysis was conducted as for the whole brain analysis. Only a subset of the contrasts listed in Table 11 were considered in the ROI analysis, namely the direct comparisons between “neighbouring” representational levels: C5′ (Explicit > Paraphrase), C6′ (Paraphrase > Inference), and C7′ (Inference > Unrelated). For illustration, activation time courses were extracted using FIR-models as in Experiment 1.

Table 12: Regions of interest in Experiment 2

ROI	Description	L/R	Brodmann Area	MNI-Coordinates	Size in mm^3
1	pMTG/STG	L	22, 21	−54, −44, 4	3328
2	pSTG, supramarginal gyrus	L	22, 39	−58, −54, 18	2840
3	mMTG	L	21	−56, −6, −10	2744
4	aSTG	L	38	−46, 22, −24	752
5	pSTG	L	22	−40, −54, 12	1976
6	pMTG, angular gyrus	L	39	−46, −70, 28	2768
7	pTL, subcortical	R	-	32, −50, 8	1312
8	Cuneus	L	18	−14, −90, 16	1320
9	Cuneus, lingual gyrus	R	17, 18	12, −82, 2	2128
10	PCC	L	31, 23	−6, −58, 20	2400
11	Lingual gyrus	L	18, 19	−10, −62, −4	744
12	pSTG, insula	R	41	44, −26, 6	1336
13	pMTG/STG	R	39	58, −68, 20	992
14	vmPFC	L/R	11	−4, 52, −18	1096
15	dmPFC	L	9	−6, 54, 36	928
16	IFG	L	47	–46, 36, −18	1848
17	IFG	L	45	-54, 30, 6	648

C.2. Results

C.2.1. Behavioral Results

The averaged behavioral results are shown in Table 13. Two ANOVAs revealed that there was a significant effect of the factor sentence condition (explicit, paraphrase, inference, unrelated, pseudowords) on response accuracies ($F_{4,48} = 9.11$, $p < .01$) as well as on reponse latencies in correct trials ($F_{4,48} = 26.13$, $p < .01$). Planned comparisons on the accuracy data were significant for the comparison of explicit and paraphrase condition ($t_{12} = 4.0$, $p < .01$), and paraphrase condition and inference condition ($t_{12} = 3.90$, $< .01$). For the response times it was found that pseudoword and explicit condition differed ($t_{12} = 6.60$, $p < .01$) as well as paraphrase and inference condition ($t_{12} = 3.17$, $p = 0.01$).

Table 13: Response times and accuracies in Experiment 2

	Response times in ms (SD)	Rel. frequency (SD)
Pseudowords (correct)	840 (133)	.97 (.05)
Explicit ("Yes")	1223 (274)	.98 (.04)
Paraphrases ("No ")	1324 (226)	.88 (.11)
Inference ("No")	1231 (238)	.97 (.05)
Unrelated ("No")	1149 (232)	.97 (.04)

C.2.2. fMRI Results

C.2.2.1. Whole Brain Analysis

The presentation of the results of the whole brain fMRI analysis is split up into two parts. First, the main contrasts for the reading and test phase are reported (C1'-C7' in Table 11). These results are supplemented by the analysis of general recognition effect (C14). Afterwards, the processing of sentence recognition statements is compared to the processing of pseudoword statements (C10'-C13').

Reading Regressor. Locations and further specifications of significant clusters concerning the reading phase of the experiment are shown in Table 14.

Reading Real Words > Pseudowords. See Figure 15 for cortical activation maps. Like in Experiment 1, reading real words as opposed to reading pseudowords resulted in extensive activation along the entire left superior temporal sulcus, from the anterior temporal poles to the posterior temporo-parietal junction, reaching into areas in the inferior parietal lobe (AnG, SMG). In the frontal cortex further clusters were found in the left IFG (BA 47), in the vmPFC (BA 11) bilaterally, and in the left dmPFC (9/10). A large activated area along the posterior midline included bilateral parietal regions (PCC) and occipital regions (cuneus). Finally, there were three clusters in right temporal lobe areas: one cluster overlapping with the insula and transverse temporal gyrus, a second area at the temporo-parietal junction (MTG, STG), and a third sub-temporal cortex region near the lateral ventricle (not illustrated in Figure 15).

Recognition task: Reading words > Reading pseudowords

Words > Pseudowords

1. STS, IFG, IPL
2. pTL
3. Cuneus, PCC
4. STG, Insula
5. MTG, STG
6. mPFC
7. mPFC

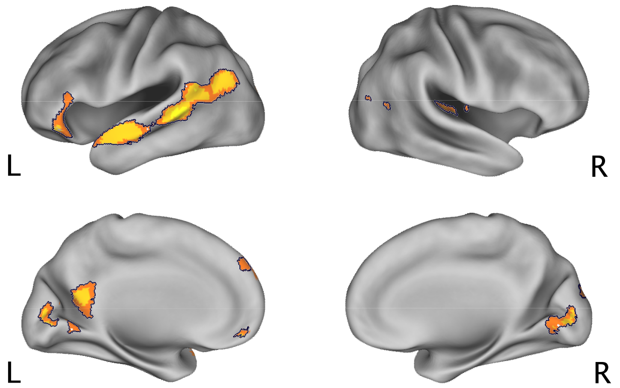


Figure 15: Significant clusters from the “Reading words > Reading pseudowords” comparison of Experiment 2 visualized on lateral and medial surfaces of inflated brain templates.

Reading phase: Explicit > Paraphrase

Explicit > Paraphrase

1. Precuneus, cuneus
2. Postcentral/precentral gyrus

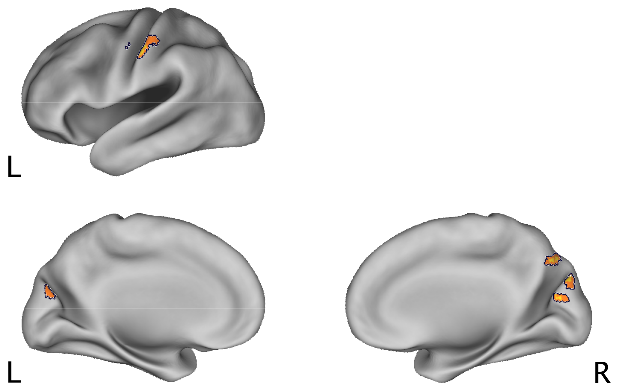


Figure 16: Significant contrasts of whole brain analysis of the reading regressor in Experiment 2. Reading explicit sentences yields greater activation than reading paraphrase sentences in one parieto-occipital cluster (precuneus and cuneus) as well as in an areas at the left central sulcus (primarily in the postcentral gyrus).

Table 14: Results from whole brain analysis of the reading phase of Experiment 2. A cluster-level threshold of $p < .1$ was applied. MNI-coordinates denote the voxel with highest z-value within the cluster.

Contrast / Region	Side	Brodmann area (BA)	Size (mm^3)	p_{corr}	z_{max}	MNI-coordinates (x, y, z)
Real words > Pseudowords						
1. Temporal lobe (MTG, STG), IFG, angular gyrus, supramarginal gyrus	L	21, 22, 39, 38, 47	26288	<.01	4.72	–54, –44, 4
2. Subgyral, pTL, lateral ventricle	R	-	1320	.03	4.66	32, −50 , 8
3. Occipital lobe, cuneus, posterior cingulate gyrus	L/R	18, 31, 17, 23, 30	12136	<.01	4.27	−14, −90, 16
4. STG, insula	R	41	1496	.02	4.16	44, −26, 6
5. MTG, STG	R	39	1608	.01	4.09	58, −68, 20
6. mPFC	L/R	11	1376	.03	3.97	−4, 52, −18
7. mPFC	L	9/10	1208	.04	3.58	−6, 54, 36
Explicit > Paraphrase						
1. Precuneus, cuneus	R/L	18, 19	5160	<.01	4.93	8, −78, 44
2. Postcentral gyrus, precentral gyrus	L	3, 2, 1, 4, 7	1544	<.01	4.30	−56, −14, 44

Comparing the sentence conditions at the time of the recognition test (recognition regressor) led to significant results for the contrast of "Inferences > Paraphrases" and "Inferences > Unrelated" statements.

Reading: Comparisons of Sentence Conditions As Figure 16 reveals, in the analysis of the main comparisons (C2′: Explicit vs. Paraphrases, C3′: Paraphrases vs. Inferences, and C4′ Inferences vs. Unrelated) for the reading regressor, significant results were only found for the contrast of explicit condition and paraphrase condition (C2′). Reading sentence endings which would subsequently be followed by explicit recognition statements produced greater activation than reading paraphrase sentence endings in bilateral midline areas in the occipital lobe (precuneus, cuneus) and in the left postcentral gyrus.

Recognition Regressor. Firstly, in Figure 17 and Table 15 the results of contrasting the recognition regressor of the sentence conditions directly (C5′, C6′, C7′) are presented together with the contrast reflecting comparisons between Hits and Correct rejections (C14). The second analysis presents which areas are more activated during the recognition of each of the sentence conditions as compared to the recognition of pseudoword test probes (Figure 18, Table 16).

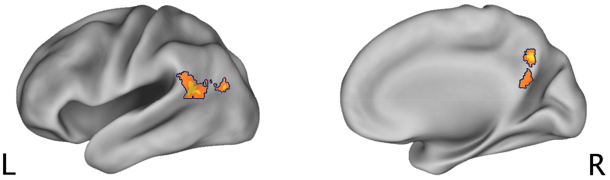
Table 15: Results from whole brain analysis of the test phase of Experiment 2. A cluster-level threshold of $p < .1$ was applied. MNI-coordinates denote the voxel with highest z-value within the cluster. *Hits:* explicit statements, *Correct rejections:* average of paraphrase, inference, and unrelated statements.

Contrast / Region	Side	Brodmann area (BA)	Size (mm3)	p_{corr}	z_{max}	MNI-coordinates (x, y, z)
Inference > Paraphrase						
1. Precuneus, cingulate gyrus	l/r	7, 31	2384	< .01	4.23	6, −64, 42
2. STG, supramarginal gyrus	l	39, 40	4552	< .01	4.14	−52, −72, 24
Inference > Unrelated						
1. Cingulate gyrus	l/r	23	1672	< .01	4.67	−8, −16, 34
2. Middle frontal gyrus	l	46	1208	.02	4.12	−42, 54, 8
3. Middle frontal gyrus, IFG	r	10	1184	.03	4.11	44, 46, 2
4. IFG	l	47, 13	1168	.03	3.76	−38, 20, −4
Hits > Correct rejections						
Inferior/Superior parietal lobe	l	40, 7	4072	<.01	4.47	−40, −48, 46
Correct rejections > Hits						
IFG	l	47	2608	<.01	4.56	−30, 20, −22

Recognition test phase: Main contrasts

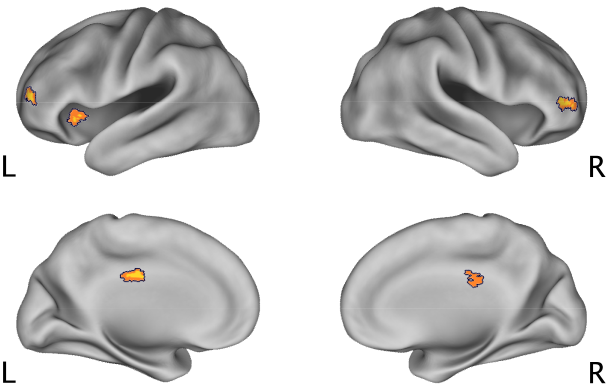
Inferences > Paraphrases

1. Precuneus, cingulate gyrus
2. STG, supramarginal gyrus



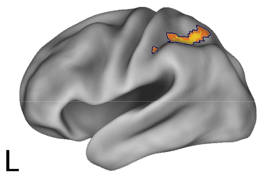
Inferences > Unrelated

1. Cingulate gyrus
2. Middle frontal gyrus
3. Middle frontal gyrus, IFG
4. IFG



Explicit > Other (Hits > CR)

Inferior/Superior parietal lobe



Other > Explicit (CR > Hits)

IFG (BA 47)

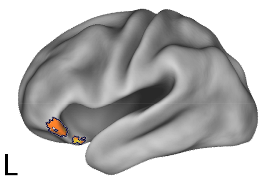


Figure 17: Visualization of the main results from the whole brain analysis of the recognition regressor in Experiment 2. The comparison between explicit and paraphrase condition did not yield significant results.

Inferences > *Paraphrases*. Two clusters were associated with more activation in inference condition as compared to the paraphrase condition: A bilateral posterior midline area (posterior cingulate gyrus, precuneus) and a region at the junction of the left STG and SMG.

Inferences > Unrelated. The processing of inference statements produced greater activation than the processing of unrelated recognition statements in the following four areas: (1) the posterior cingulate gyrus, (2) the left frontopolar area 10, (3) the right frontopolar area 10, and (4) the left IFG (BA 47).

Recognition effects: Hits versus Correct rejections. More activation in explicit trials (Hits) than in the average of paraphrase, inference, and unrelated trials (Correct rejections) was associated with a region in the left parietal lobe. This cluster covered parts of the inferior parietal lobe (BA 40) as well as parts of the superior parietal lobe (BA 7). This region around the inferior parietal sulcus has been found in other studies comparing the recollection of old items and the rejection of new items (Kahn, Davachi, & Wagner, 2004; Wagner, Shannon, Kahn, & Buckner, 2005). Trials that required a no-response (paraphrase, inference, and unrelated trials) produced more activity than trials that required a yes-response (explicit) elicited more activation in the left IFG (BA 47).

Pseudoword condition as baseline. Explicit statements produced more activation than pseudoword statements (C10′) in one cluster comprising parts of the left IFG and MFG. The processing of paraphrase statements in comparison the pseudowords (C11′) was also associated with greater activation than the processing of pseudowords in the left lateral prefrontal cortex, reaching from ventral aspects (BA 47) to the dorsal BA 6. Two further clusters were located in the left posterior MTG/STG and in posterior portions of the dmPFC (BA 8, 6). Inference statements led to more activation than pseudowords (C12′) in the left lateral PFC and in left temporal areas along the STS. The temporal activation consisted of two clusters, one reaching from the temporal pole to the posterior MTG/STG and a smaller, distinct cluster at the border to the angular gyrus. Finally, comparing the unrelated condition with pseudowords (C13′) resulted in one large cluster that comprised the left lateral PFC and left temporal areas along the STS. The pattern of activity was similar to that of the comparison of inference condition and pseudowords, where the prefrontal activation extended more dorsally.

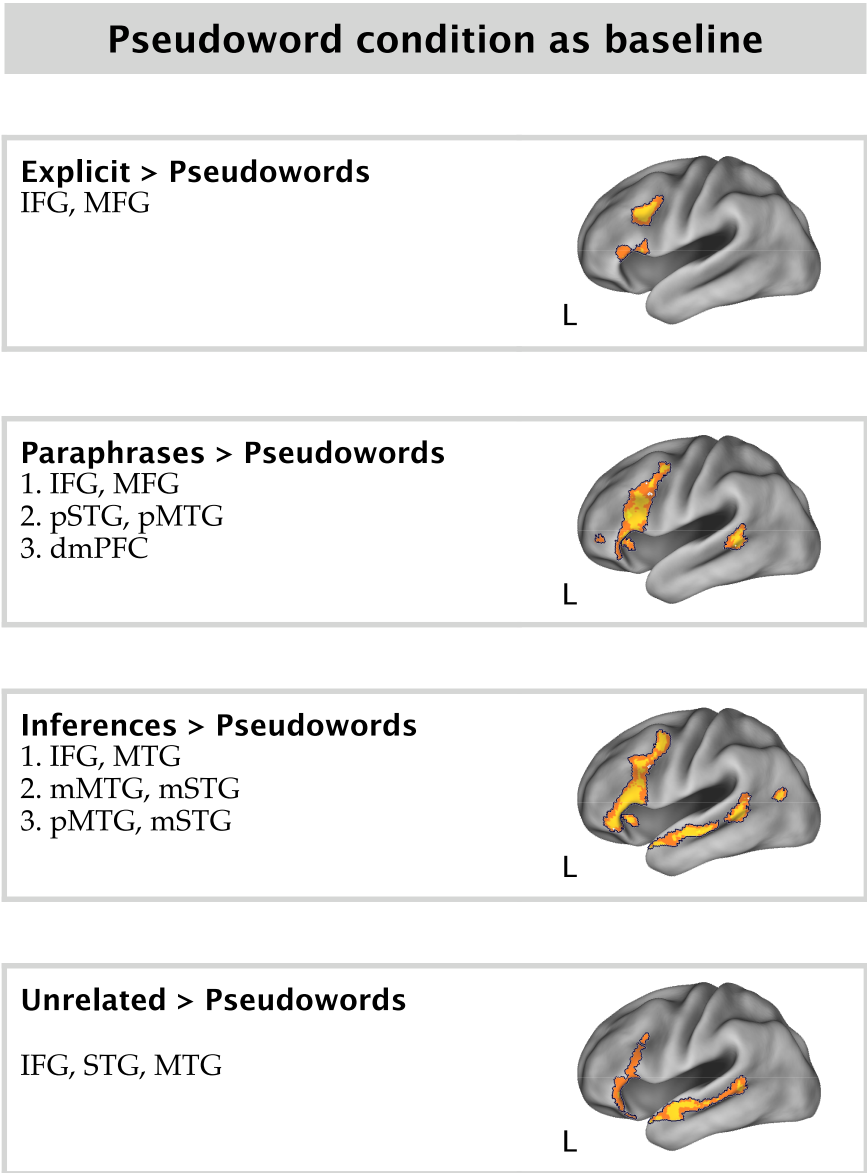


Figure 18: Comparisons against the pseudoword condition as a common baseline in the whole brain analysis of the recognition regressor (Experiment 2). No significant clusters were found in the right hemisphere or in the medial aspects of the left hemisphere.

Table 16: Pseudoword condition as common baseline. Results refer to the whole brain analysis of Experiment 2.

Contrast / Region	Side	Brodmann area (BA)	Size (mm^3)	p_{corr}	z_{max}	MNI-coordinates (x, y, z)
Explicit > Pseudowords						
IFG, MFG	L	47, 45, 44, 9	7088	<.01	4.86	−36, 16, 32
Paraphrase > Pseudowords						
1. IFG, MFG	L	47, 45, 46, 9, 44, 6	14344	<.01	4.95	−52, 20, 14
2. pSTG, pMTG	L	22, 21	2168	<.01	4.63	−60, −44, 6
3. dmPFC	L	6, 8	1464	.02	4.40	−4, 10, 58
Inference > Pseudowords						
1. IFG, MTG	L	47, 45, 46, 9, 44, 6	17360	<.01	5.19	−36, 6, 44
2. mMTG, mSTG	L	21, 38, 22	4704	<.01	4.93	−54, −16, −10
3. pMTG, mSTG	L	22, 39,21	4904	<.01	4.32	−50, −44, 2
Unrelated > Pseudowords						
IFG, STG, MTG	L	21, 38, 47, 45, 46	12344	<.01	4.55	−44, 20, −22

C.2.2.2. Regions of Interest Analysis

The statistical analysis in the regions of interest analysis of differences between conditions was based on the predefined regressors for the reading phase and the test phase of the experiment. Within each region of interest, difference contrasts between (C5′) Paraphrase versus Explicit, (C6′) Inferences versus Paraphrase, and (C7′) Unrelated versus Inference statements were calculated for the regressor representing the reading of the last six words of the context sentence and for the regressor representing the execution of the recognition test. All contrasts were calculated in both directions, i.e. “Paraphrase *greater than* Explicit” and “Explicit *greater than* Paraphrase” and so on. The results are shown in Table 17.

Table 17: Results from the regions of interest analysis of Experiment 2. Reported are only regions in which at least one contrast was associated with *p*corr < .1.

ROI	Area	Phase	Contrast	Statistics	
				t	p_{corr}
2	l-pSTG	Test	Inference > Paraphrase	7.03	<.01
		Test	Inference > Paraphrase	7.03	<.01
3	l-mMTG	Test	Inference > Paraphrase	3.90	.02
		Reading	Inference > Paraphrase	3.13	.07
6	l-pMTG	Test	Inference > Paraphrase	4.40	<.01
10	l-pCC	Test	Inference > Paraphrase	3.19	.06
13	r-pMTG/STG	Test	Inference > Paraphrase	3.70	.03
14	vmPFC	Reading	Unrelated > Inference	3.15	.07
		Test	Inference > Paraphrase	4.51	<.01
15	dmPFC	Test	Inference > Paraphrase	3.18	.07
16	l-IFG (BA 47)	Test	Paraphrase > Explicit	2.98	.09
		Test	Inference > Paraphrase	3.29	.05

Of these comparisons the contrast C6′ "Inference > Paraphrase" produced significant results in the test phase of the experiment in the left posterior STG (ROI 2), in the left posterior MTG (ROI 6), in the right posterior STG/MTG (ROI 13), and in the left ventro-medial PFC (ROI 14). Furthermore, almost significant results ($p < .1$) were obtained in the posterior cingulate gyrus (ROI 10), in the dorso-medial PFC (ROI 15), and in the left ventro-lateral PFC (ROI 16). The reversed contrast "Paraphrase > Inference" did not yield significant results. No significant differences were found in the comparisons of explicit and paraphrase statements (in both directions), neither for the reading regressor (C2′), nor for the test regressor (C5′). In the analysis of the reading regressor two contrasts were associated with a *p*-value below .1 (corrected for multiple comparisons). The left anterior STG (ROI 4) was more activated by inferences than by paraphrases (C3′) and in the ventro-medial PFC (ROI 14) unrelated statements led to greater activity than inferences (C4′). Estimated activation time courses of the different conditions were included to illustrate and supplement the inferential statistics presented above.

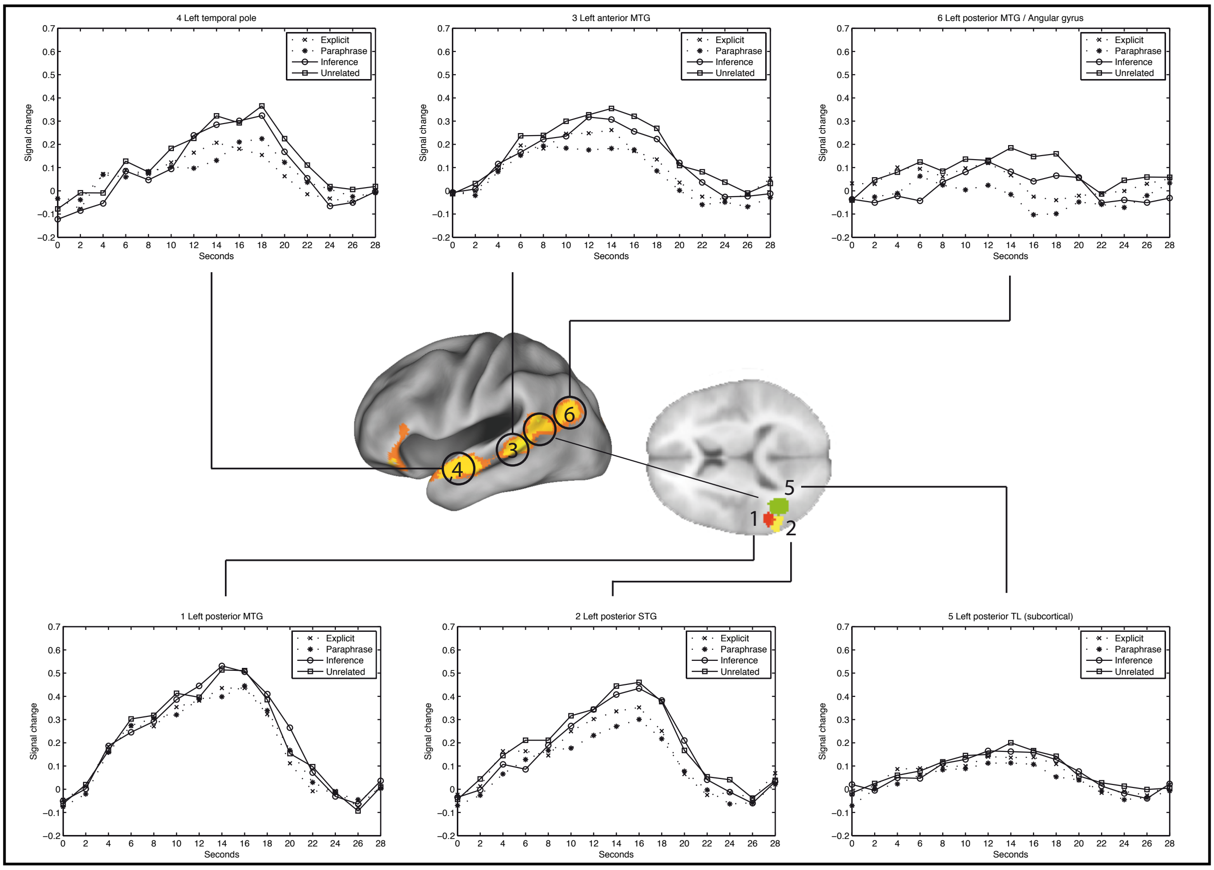


Figure 19: Activation timecourses in left temporal regions of interest. Time course diagrams show the estimated signal change in each ROI for the four sentence conditions from the onset of the headline to the end of the trial. See Table 12 for specifications of the ROIs and Table 17 for details of the statistical analysis.

Left temporal cortex. In most left temporal regions (ROI 1-5), except for the posterior MTG and the angular gyrus (ROI 6), the time courses of unrelated and inferences trials on the one hand, and on the other hand paraphrase and explicit trials followed relatively similar patterns (Figure 19). In ROI 1-5 unrelated and inferences trials tended to be more activated than paraphrase and explicit trials throughout the whole trial length. Additionally, 14 to 16 seconds after trial onset—when the BOLD response was maximal in most regions—the unrelated condition reached slightly higher activations than the inference condition, and in explicit trials activations were higher than in paraphrase trials. In ROI 4 (left temporal pole) unrelated, inference, and paraphrase statements led to maximal activation only after 18 seconds. Overall, activations were lower in the posterior regions ROI 5 and ROI 6 than in the more anterior temporal regions. In ROI 6 deactivation of the paraphrase condition occurred between 16 and 18 seconds after trial onset.

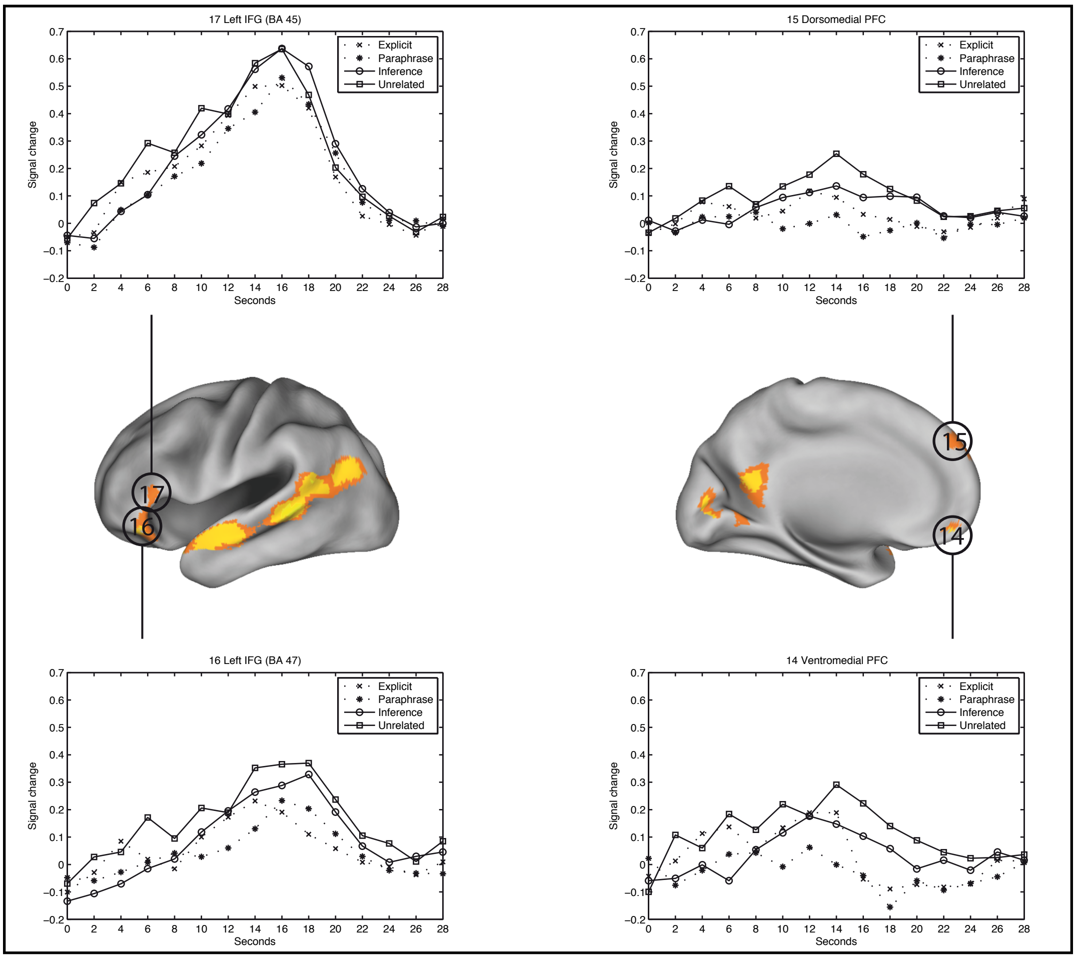


Figure 20: Activation timecourses in medial and lateral prefrontal regions of interest. Time course diagrams show the estimated signal change in each ROI for the four sentence conditions from the onset of the headline to the end of the trial. See Table 12 for specifications of the ROIs and Table 17 for details of the statistical analysis.

Prefrontal cortex. The overall pattern of activation timecourses in the prefrontal regions of interest again revealed higher maximum values for unrelated and inference condition than for explicit and paraphrase condition (Figure 20). Moreover, activations were higher in the unrelated condition than in the inference condition, and activations in the explicit condition were either higher than in the paraphrase condition (ROI 14, 15), or within a comparable range (ROI 16, 17).

ROI 16, located in the ventral aspect of the left IFG (BA 47), showed a similar resonse pattern as ROI 4 in the left anterior temporal lobe. Peak amplitudes were reached 18 seconds after trial onset for unrelated and inference trials, while explicit and paraphrase trials developed maximal activity slightly earlier. Highest responses for all conditions and across all regions of interest were estimated in ROI 17 in the left IFG (BA 45) after 16 seconds.

Right temporal and parietal cortex. In the right posterior MTG (ROI 13), at the boundary of temporal, parietal and occipital cortex, the activation pattern was similar to that of the contralateral area in the left hemisphere, corresponding to ROI 6 (Figure 21). Timecourses in ROI 13 tended to follow a bimodal run. While maximal activations were developed after 12-14 seconds, there was a second, lower maximum after 4-6 seconds. Again, in the time window around the maxima, highest BOLD responses were produced by the unrelated condition, followed by inference and explicit condition, with the paraphrase condition causing the lowest signals.

Activation timecourses in the right insula cortex (ROI 12) and in the subcortical area near the lateral ventricle (ROI 7) did not show a clear peak, and did not differentiate between conditions.

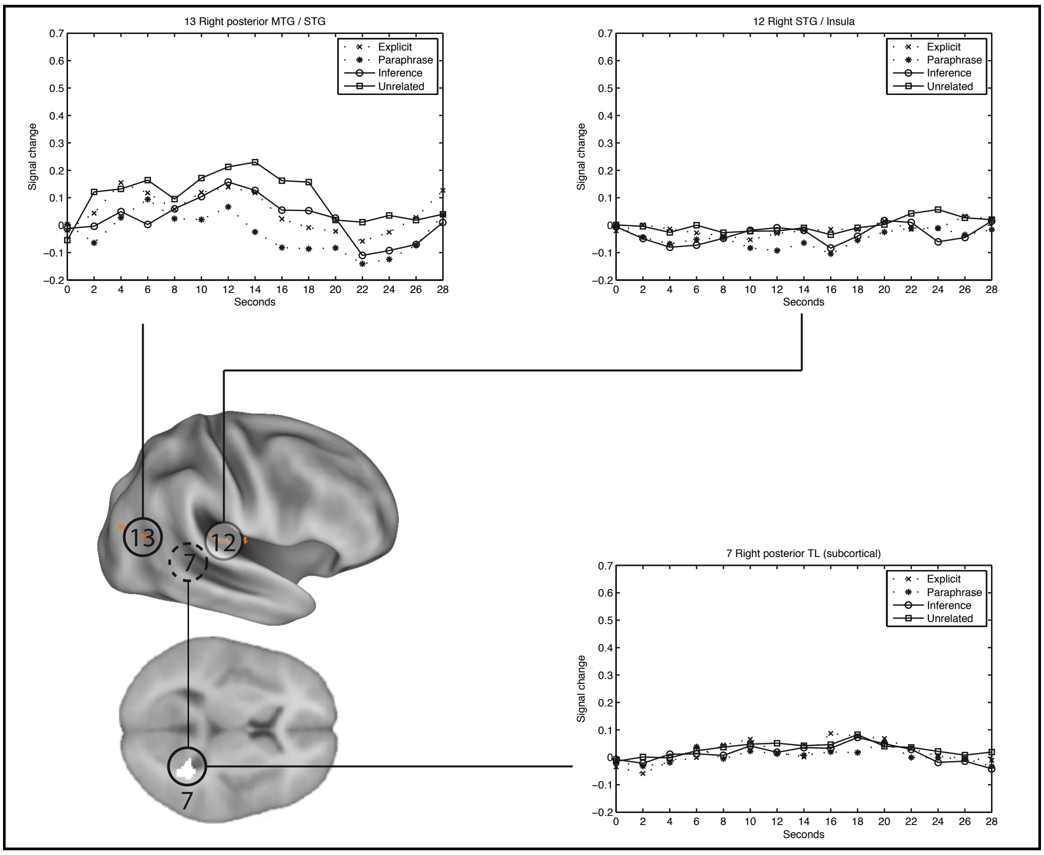


Figure 21: Activation timecourses in right temporal and parietal regions of interest. Time course diagrams show the estimated signal change in each ROI for the four sentence conditions from the onset of the headline to the end of the trial. See Table 12 for specifications of the ROIs and Table 17 for details of the statistical analysis.

Posterior midline areas. Comparable BOLD responses in the secondary visual cortex areas (left and right cuneus, ROI 8 and 9) were observed for all conditions (Figure

22). Time courses showed a bimodal run with a first maximum after 4 seconds, followed by a short phase of deactivation at 6-12 seconds, rising then to the absolute maximum at 16-18 seconds. This pattern was also seen in the left lingual gyrus (ROI 11).

In the posterior cingulate gyrus (ROI 10) global maxima were comparatively low, and occurred after 4-6 seconds in all conditions except for the unrelated condition (peak after 14 seconds). There was a noticeable deactivation of paraphrases that developed after 8 seconds.

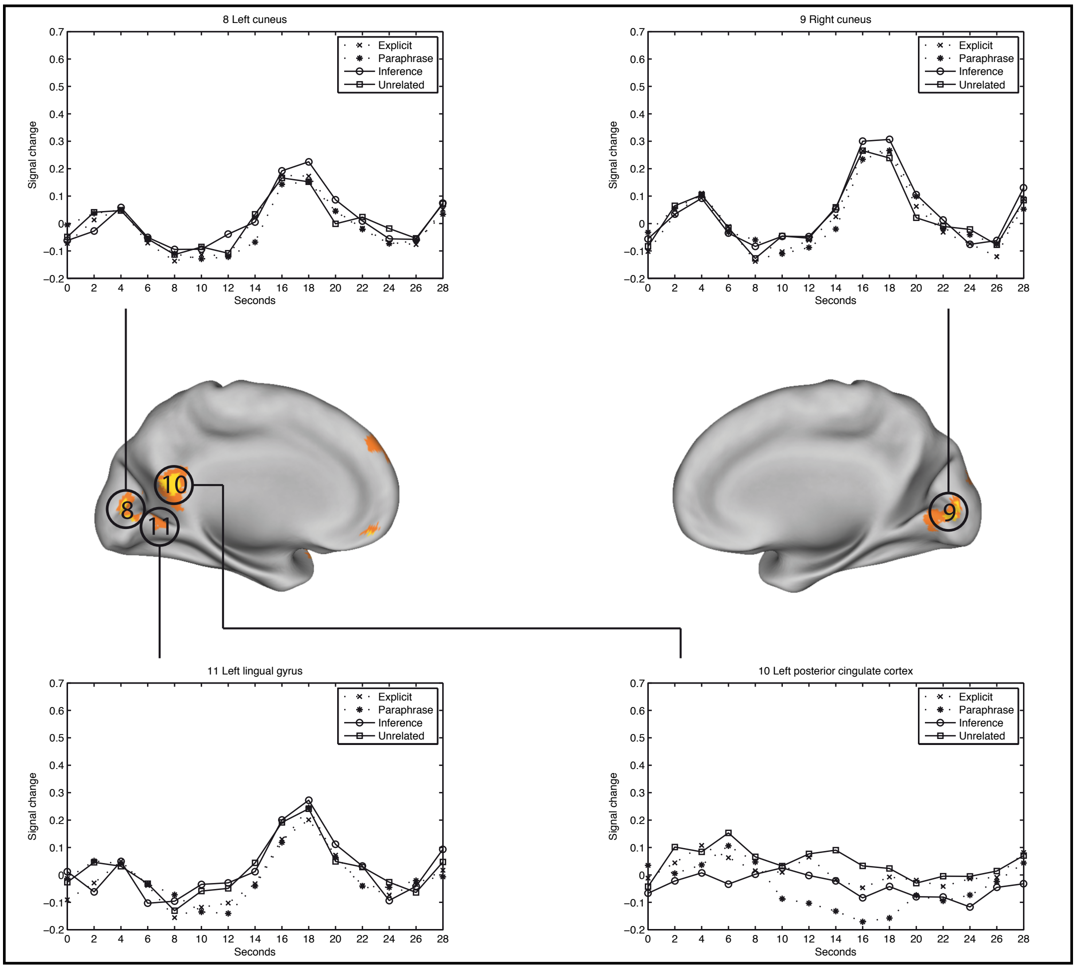


Figure 22: Activation timecourses in posterior-medial and occipital regions of interest. Time course diagrams show the estimated signal change in each ROI for the four sentence conditions from the onset of the headline to the end of the trial. See Table 12 for specifications of the ROIs and Table 17 for details of the statistical analysis.

C.3. Discussion

In this experiment participants read single context sentences and immediately afterwards received short statements for recognition. These recognition statements matched words from the context sentence either at all levels of representation (explicit condition), at propositional and situational levels (paraphrase condition), only at the situational level (inference condition), or they were unrelated to the context sentence (unrelated condition).
The behavioral results showed that paraphrase statements produced significantly more false positive responses and slower response times than inferences. Presumably, paraphrase statements were more difficult to reject because they shared the propositional representation with the context sentence. The identical meanings forced the comparison mechanism to focus on the surface level representations and to retrieve and compare the exact wording of the context sentence. It can be assumed that the rejection of unrelated statements was easier (faster and more accurate) because the lack of coherence between test statement and context sentence additionally pointed to a no-response. Accordingly, responses to inference statements, which match the context sentence at the situation level, were slower than responses to unrelated statements. The behavioral results thus provide evidence for the validity of the claim that the proposed comparison process takes all three levels of representations into account, although the recognition instruction focusses the readers´ attention on the surface level. Consistent with this interpretation, Reder (1982) demonstrated that it is usually more efficient to judge the plausibility of a statement than to retrieve the verbatim context.

In the remainder of this chapter, the most important findings from the whole brain fMRI analysis will be discussed. Afterwards, the inference condition will be considered in more detail, and some preliminary conclusions will be drawn.

Reading context sentences (in comparison to reading pseudowords) was associated with a pattern of cortical activity in widely distributed brain areas that was similar to the network identified in Experiment 1. This included regions along the left STS, in the left IFG, vmPFC, and dmPFC. Moreover, activity was also found in right TL areas, and in medial occipital areas (Figure 15, Table 14). The occipital activations, which were not found in Experiment 1, might have been caused by more intense visual processing of the context sentences due to the focussing on the surface representation by the recognition instruction.

The significant clusters found in the comparison of explicit and paraphrase condition during the reading phase of the experiment were unexpected. Reading explicit sentences produced significantly more activity than reading paraphrase sentences in the left postcentral gyrus and a medial parieto-occipital cluster (precuneus, cuneus). In the context of studies that are related to language processing in the widest sense, activation in sensorimotor areas like the bilateral postcentral gyri has been found in comparisons of lexical decision task versus reading aloud (Carreiras, Mechelli, Estevez, & Price, 2007) and general typing finger movements (Gordon, Lee, Flament, Ugurbil, & Ebner, 1998). Precuneus and cuneus have been found to be involved in, for instance, narrative comprehension (Xu, Kemeny, Park, Frattali, & Braun, 2005) and reasoning tasks (Ruff et al., 2003). It is difficult to figure out why exactly these areas were activated in this contrast. The above functional characterizations are "reverse inferences" (Poldrack, 2006) which seem to be poorly justified in this case.

There are at least two possible reasons for this unexpected result. As the sentence materials were constructed on the basis of the respective explicit sentence versions, paraphrase sentences often contained less frequently used words or expressions. Although care was taken not to construct noticeably uncommon wordings, paraphrase sentences nevertheless might have been discriminable. As the instruction focussed attention on the exact wording of the sentences, participants of Experiment 2 might have been more sensitive for these effects than participants from Experiment 1 in which no similar effects were observed. Indeed, two participants of Experiment 2 mentioned in post-experiment interviews that they sometimes felt they were able to predict the upcoming recognition statement. Though, removing the data of these participants from the analysis did not change the overall results pattern—speaking against a major influence of this aspect. Furthermore, if participants were able to predict the subsequently required response, there should not have been significant differences in the behavioral data. Additional evidence that this result might not be related to properties of the sentence material comes from the regions of interest analysis. There was no region in which the explicit condition produced significantly higher activation than the paraphrase condition, and the time course data in the parieto-occipital regions of interest did not indicate noticeable differences between the explicit and paraphrase condition (Fig. 22; ROI 8, 9, 11).

Another possible explanation for this unexpected finding is that it might simply represent a methodological artifact due to the nature of the underlying GLM-based analysis method. As the regressors for the reading and the test phase of this experiment are not completely independent, explained variance in the model cannot always unambiguously be attributed to one regressor or the other. Thus, in principle, activity might "spill over" from the regressor representing the recognition test to the regressor representing the reading phase by chance. Against this explanation speaks that the two regressors for the respective reading and the test phase were separated by 1.7 seconds, and were thus clearly distinct.

In all accounts, the fact that participants might have been able to predict the upcoming statements based on properties of the context sentences constitutes a serious shortcoming of the present study. Although, for the reasons mentioned above, it seems unlikely that participants actually used prediction strategies in the recognition task, an influence of such strategies on the results can not be ruled out completely. Therefore, the results have to be interpreted with caution.

In contrast to the first experiment with the verification task, in Experiment 2, paraphrase and inference statements (as well as unrelated statements) required a no-response. More activity in inference than in paraphrase trials was assumed to reflect the comparison of the representations of test statement and context sentence on the situational level with the rejection component cancelled out by the subtraction. Significant clusters were found at the left temporo-parietal junction (STG, SMG) and in the posterior medial cortex (precuneus, PCC). Areas at the left posterior STG/MTG and SMG have been associated with several aspect of language processing, including phonological and semantical processing. Possibly the area is involved in mapping phonologically-based representations to semantic concepts (semantic activation; Jung-Beeman, 2005). Mid-line parietal cortex areas (cingulate gyrus, precuneus) are supposed to be involved in memory retrieval processes (Bottini et al., 1994; Ferstl & von Cramon, 2002). A possible explanation could be that, in the inference condition, the match of the situation model representations of context sentence and recognition statement caused interference with the required rejection. Therefore, a reinstatement process, supported by the PCC, may have been executed to check the representation of the context sentence. Furthermore, participants might have attempted to actively rehearse the context sentence, a process which would likely occur with contributions of pSTG/SMG.

No significant clusters were found in the reversed comparison, paraphrases > inferences, despite the behavioral results demonstrated that in the paraphrase condition responses were significantly slower and less accurate than in the inference condition.

The comparison of inference and unrelated statements (Inference > Unrelated) was not related to specific a priori hypotheses. The evaluation that inference statements are plausible within the context of the study sentence, whereas the unrelated statements are not plausible, was associated with significant clusters in the bilateral frontopolar cortices (BA 10), the left IFG (BA 47), and the mid cingulate cortex (BA 23). Although this contrast did not yield significant results in Experiment 1, it seems plausible that the evaluation process for the unrelated statements was less intense than for inference statements with the recognition task. As has been argued above, matching representations at the propositional or situational level can be interpreted as a source of distraction because the semantic or contextual fit interferes with the required no-response. In this sense, unrelated statements would not produce interference because they do not match the representation of the context sentence on any level. Areas showing greater activation for inference than for unrelated statements might thus reflect interference based on the evaluation of the plausibility of the inference statement. Correspondingly, response times were descriptively larger in the inference condition than in the unrelated condition. Activation of the frontopolar regions has been related to executive functions, like control processes during memory retrieval (Buckner, 2003). It is possible that these regions contribute to the decision making process during the rejection of inference statements more strongly than during the comparatively easier rejection of unrelated statements. The triangular part of the left IFG (BA 47) presumably contributes to the more intense semantic processing of inference statements (Jung-Beeman, 2005; Wagner et al., 2001). Activity in the cingulate cortex could be associated with higher demands on memory retrieval processes during the evaluation of the inference statements (Cabeza & Nyberg, 2000b).
To see if the results from the statement recognition procedure are compatible with the more general recognition memory literature, hits and correct rejections were analyzed in two contrasts. Yes-responses to explicit statements (Hits) yielded more activity than no-responses to paraphrase, inference, and unrelated statements

(Correct rejections) in regions of the left parietal lobe. Correct rejections, on the other hand, in comparison to Hits were found to engage the left IFG (BA 47). Both of these regions have been shown to be modulated by retrieval success (McDermott, Jones, Petersen, Lageman, & Roediger, 2000). The IPL is consistently found to reflect successful retrieval (Buckner & Wheeler, 2001; Henson, Hornberger, & Rugg, 2005; Kahn et al., 2004). The present study confirmed this finding although recognized items (the recognition statements) were embedded in the sentence context in the study phase. This is different from most general recognition memory experiments, in which word lists are usually presented without context. The activity in the ventral IFG (BA 47), yielded in the contrast of Correct rejections > Hits, most likely reflects increased semantic processing for the rejected statements.

C.3.1. Inference Processing in a Sentence Recognition Task

The most reliable findings of this experiment are represented in the comparisons of inference and paraphrase condition in the whole brain analysis and in the regions of interest analysis. In the whole brain analysis, inference statements yielded greater activation of the left temporo-parietal junction area and in the posterior medial cortex (precuneus, PCC; Figure 17, Table 15). This was confirmed by the same contrast in the corresponding regions of interests, which was significant for ROI 2 in pSTG/SMG and almost significant (p = .06) for ROI 10 in the PCC (Figures 19 & 22, Table 17).

As noted above, pSTG and SMG are supposed to contribute to semantic activation or general phonological processing (verbal working memory). Activation of the PCC is often related to memory retrieval processes. However, note that the relative greater activation of inference statements in the PCC rather appears to be caused by greater deactivation of paraphrase statements (see the time course plot for ROI 10 in Figure 22). There is evidence that the PCC and precuneus are active in awake human participants when no attention-demanding task is performed (Raichle et al., 2001). This “default state” is thought to reflect the gathering and evaluation of external and internal information during rest. Relative deactivations of the PCC and precuneus have been found frequently in a variety of cognitive tasks, and it is supposed that attenuation of the default activity reflects increasing focused attention on the task at hand (Raichle & Snyder, 2007; Raichle et al., 2001). By this notion, greater deactivation of paraphrase statements than inference statements might indicate that paraphrase statements required more processing resources in some aspects, which would be compatible with the slower response times and lower response accuracies in comparison to the inference condition. This interpretation is highly speculative though and the current experiment provides far more evidence for greater activation in the inference than in the paraphrase condition in several other areas.

Apart from the regions discussed to far, inference statements additionally yielded significantly higher activation than paraphrase statements in the left temporal lobe (ROI 3/mMTG, ROI 6/pMTG), in ROI 13 in the right pMTG/STG, and in ROI 14, located in the vmPFC. Almost significant ($p < .1$) results were obtained in the dmPFC

(ROI 15) and in the left ventral IFG (ROI 16). This is particularly interesting because these regions were identified in Experiment 1 as being central for inferencing. Table 10 and Figures 10 & 11 show the corresponding results of Experiment 1, in which the respective contrast was also significant in the ventral IFG (ROI 2 of Exp. 1) and in the dmPFC (ROI 9 and 10 of Exp. 1). In temporal areas, significantly higher activity for inference statements than for paraphrase statements was obtained in the left mid and posterior temporal lobe regions (ROI 3, 2, 6; see Table 17). Furthermore, the inspection of the activation time courses confirms these observations in most regions of interest (Figures 19-22).

Thus, many of the regions that have been described to constitute the extended language network (Ferstl et al., 2007), including left IFG, dmPFC, PCC, and temporal lobe structures, were more active in the inference than in the paraphrase condition. Presumably, this indicates that inference statements tapped into several language processes more intensely than paraphrases. This is noteworthy because the recognition task does not require a detailed analysis of the meaning of the inference statements or an evaluation of the contextual fit. A system, tuned to perform this task most efficiently, would be expected to just evaluate if the surface level representations of the recognition statements matches those of the context sentences. Instead, in human language comprehension new incoming words always need to be mapped *conceptually* to the previously received text. The results of the present study can be interpreted such that, even when the processing goal is recognition, this reinstatement does not only occur at the textbase level but involves inferences from the situational level.

An alternative, more parsimonious explanation for the greater activity during the recognition of inference statements as compared to paraphrase statements in the distributed extended language network could claim that all of these regions simply contribute to the memory search process, with increasing activity reflecting decreasing similarity of context sentence and recognition statement. However, this interpretation seems unlikely for two reasons. First, it ignores well-established findings of functional specializations of cortical areas, i.e. the principle of functional segregation. Second, this explanation would predict slower response times to unrelated statements than to inference statements, which is not supported by the results. Descriptively, responses to inference statements were even slower than responses to unrelated statements.

C.3.2. Summary and Intermediate Conclusion

The regions of interest analysis revealed that inference statements were associated with higher cortical activation than paraphrase statements in most of the functionally defined ROI, including the left IFG, left MTG/STG, dmPFC, and PCC. Additionally, significantly higher activation in the left posterior temporal lobe and in the PCC was also found in the whole brain analysis. Although activity in the dmPFC and left IFG, which were associated with inference processing in Experiment 1, was higher for the inference than for the unrelated condition in the ROI analysis, this effect was not as pronounced as in Experiment 1, and there was no effect in the whole brain analysis. Taken together, these results indicate that the recognition instruction indeed attenuated inference generation processes on the situational level. The finding that in the dmPFC and also in the ventral IFG the processing of inference statements was associated with greater activity than the processing of paraphrase statements can be interpreted as evidence for situation level processing. These activations are thought to reflect that the participants routinely evaluate the meaning of the recognition statement in the context of the study sentence, although—with respect to the recognition task—this evaluation is dysfunctional for the performance.

D. Comparison of Experiment 1 and Experiment 2

So far, it was investigated how verification and recognition task instructions affected the processing of the study sentences and test statements in the different conditions within each experiment. In this part, effects of the tasks are assessed directly as between-subjects factor to explore which brain areas reflect global differences and commonalities of the cognitive processes engaged during the two tasks.

Participants in the two studies presumably were engaged in different processing strategies, induced by the respective tasks or goals. Such differences would not be observable by considering the fMRI results of each experiment on its own because activations common to all conditions in each experiments are likely to be cancelled out by the subtraction procedure. To explore task-specific processing, explicit, paraphrase, inference, and unrelated condition were combined, and it was tested in which brain regions the processing of these conditions produced more or less activation depending on whether the task was to verify or recognize test statements. Both, the reading phase and the test phase regressors were analyzed.

The analysis of the reading phase regressor was supposed to reflect differences in the encoding of the context sentences. In a recent study (Chow, Kaup, Raabe, & Greenlee, 2008) it was found that if participants were instructed to try to predict upcoming events during the reading of passages, activity increased in two regions within the left IFG and in a third area in the left anterior prefrontal cortex (aPFC, BA 9), slightly lateral to the dmPFC region found in the current experiments. Moreover, the authors demonstrated that different reading instructions not only modulated the total activity of the dorsal IFG, ventral IFG, and aPFC but also the connectivity between these and other regions (e.g., the pSTS). Although participants in Experiment 1 were not instructed to actively utilize prediction strategies, the underlying mechanisms for prediction and verification are likely to be similarly based on situation level processing. Thus, higher involvement of the left IFG and maybe in the dmPFC could also be predicted for reading in the verification experiment as compared to the recognition experiment in which situation level processing should not be prominent. The comparisons against the pseudoword condition in Experiment 1 (Figure 10) and Experiment 2 (Figure 18) suggest that the right aMTG, which was significantly activated in Experiment 1 but not in Experiment 2, might also contribute to a more elaborative reading in the presence of the verification task.

The region that was most prominently associated with inference processing in Experiment 1 was the dmPFC. However, the results of Experiment 2 showed that this region was also differentially activated when the participants were occupied with a recognition task. Activity in the dmPFC was found during reading (Words > Pseudowords, Figure 15), and the regions of interest analysis demonstrated significant differences between the inference and the paraphrase condition (Table 17) in Experiment 2. Directly comparing the processing of verification statements and recognition statements would thus reveal if the dmPFC was reliably more activated in Experiment 1.

In addition to investigating differences between the cortical networks involved in the two experiments, conjunction analyses were conducted to characterize which brain regions were commonly activated during verification *and* recognition. In the comparisons of reading real words versus reading pseudowords, regions in the left lateral and medial frontal cortices, the left temporal lobe, and in PCC were activated in both experiments (Figures 7, 15). Other regions were only found in one of the studies, e.g. the right aMTG in the verification experiment and medial occipital areas in the recognition experiment. Similarly, regions in the left lateral frontal cortex and in the left temporal lobe were activated in most contrasts concerning the test phases (Figures 10, 18). But whereas in Experiment 1 right aMTG and dmPFC activation was found in the contrasts with inference and unrelated statements, Experiment 2 appearently yielded stronger activations in the left dlPFC and in premotor regions. The conjunction analyses were performed to establish the core networks that supported language processing in both experiments.

Note that in this chapter all comparisons concerning the test phases of the two experiments are confounded with the different ratios for yes- and no-responses. In the verification experiment this ration was 3:2 ("yes" to explicit, paraphrase, and inference versus "no" to unrelated statements and fillers). Accordingly, in the recognition experiment the ratio was 2:3. Despite this confound, it seems justifiable to evaluate these comparisons in an exploratory manner, since in both experiments contrasts between conditions with different required responses did not yield significant results (Experiment 1: "Unrelated > Inference"; Experiment 2: "Paraphrases > Explicit").

D.1. Method

A conjunction analysis (Friston, Penny, & Glaser, 2005) was performed to identify areas that were commonly activated during reading in Experiment 1 and Experiment 2. Furthermore, the resulting activation map served as mask in the direct comparisons between the experiments. For the conjunction, the results of the random effect analysis concerning the "Reading words > Reading pseudowords" contrasts of both experiments were combined via a logical "AND" function in SPM2. An uncorrected p-value of $p = .002$ on voxel-level and a corrected p-value of $p < .05$ on cluster level were chosen as thresholds. The slightly more liberal voxel-level threshold was used to include a cluster in the PCC that was not significant at $p = .001$. The resulting statistical map (see Figure 23) was used assumed to reflect the core regions for language comprehension processes.

A second conjunction analysis was performed to find areas that were commonly activated in the verification and recognition processes at the time of the test. Therefore, for each experiment, regressors were created that reflected the processing of explicit, paraphrase, inference, or unrelated statements in comparison to the pseudoword task ("Real test words > Pseudoword test probes"). The conjunction of the resulting statistical maps was computed with an uncorrected p-value of $p = .001$ on voxel-level and a corrected p-value of $p < .05$ on cluster level.

Two-sample *t*-tests were calculated on the basis of the combined regressors for explicit, paraphrase, inferences, and unrelated condition, with the experimental task (Verification versus Recognition) as between-subject factor. To restrict the statistical analysis to the core language processing regions the statistical map displayed in Figure 23 was used as a mask. As thresholds an uncorrected p-value of $p = .001$ on voxel-level and a corrected p-value of $p < .1$ on cluster level were applied. Comparisons were made between for the reading phase of the experiment and for the test phase.

D.2. Results

For the reading phases of the experiments the conjunction analysis resulted the network of areas depicted in Figure 23.

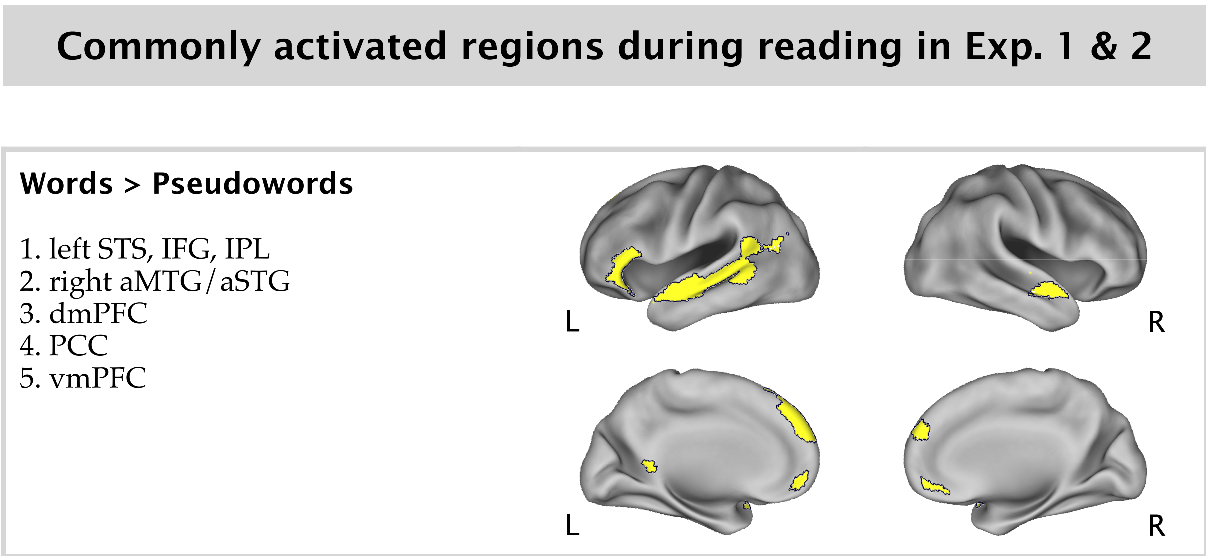
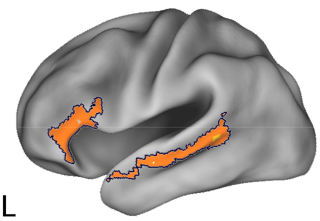


Figure 23: Results from the conjunction analysis of the reading phases of Experiment 1 and 2. Visualized are five clusters that were active in the conjunction analysis of "Reading words > Reading pseudowords" from both experiments.

Significant activations were found in peri-sylvian areas (IFG, STS, IPL), in dmPFC, vmPFC, and PCC, as well as in the right aMTG. In the second conjunction analysis for the test phases, verification and recognition tasks were found to elicit common activation in the left IFG and along the left STS (Figure #, Panel A).

Common and distinct activations during test in Exp.1& 2

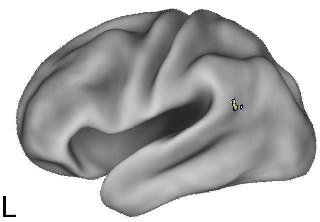
A. Verification and recognition tests commonly activate left IFG and STS



Left IFG and left STS

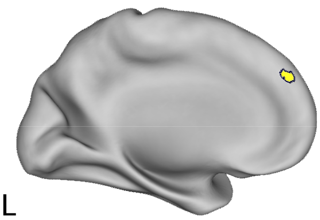
BA	mm³	p_{corr}	z_{max}	x, y, z
47, 45, 44	6144	< .001	4.38	−48, 28, −8
21, 22, 38	5120	< .001	4.22	−56, −44, 4

B. Verification activates left pSTG and dmPFC more than recognition



Left pSTG / SMG

BA	mm³	p_{corr}	z_{max}	x, y, z
39	648	.07	4.32	-58, -60, 24



Left dmPFC

BA	mm³	p_{corr}	z_{max}	x, y, z
9	496	.08	3.72	-4, 52, 32

Figure 24: Direct comparison of verification and recognition tasks. All responses to explicit, paraphrase, inference, and unrelated condition were combined in each experiment. Panel A refers to the conjunction of the verification and recognition test. The left IFG and TL were commonly activated in both tasks. Panel B shows that verifying statements leads to higher activation than recognizing statements in the left dmPFC and the left pSTG.

The subtraction analysis revealed that the two tasks did not yield significant differences with respect to the reading regressor. During reading, no area responded more to the verification instruction than to the recognition instruction, and vice versa. For the test phase of the experiments, two regions were associated with *p*-values below $p < .1$ in the contrast "Verifying statements > Recognizing statements" (Figure #, panel B). The first cluster was located in the left pSTG, and the second cluster fell into the left dmPFC. There were no regions showing higher responses to recognizing statements than to verifying statements.

D.3. Discussion

The conjunction analysis revealed that during the reading of the context sentences brain areas in lateral and medial cortices of the left and the right hemispheres were activated. The results were compatible with the regions which are often found in imaging studies on language processing ("extended language network", Ferstl et al., 2007). In the present experiments, IFG, STS, and IPL were activated in the left hemisphere during reading under both tasks. Furthermore, common activation was also found in the dmPFC, vmPFC, PCC, and in the right aMTG. The cluster in the right aMTG, which was found in the conjunction, was not significantly activated in Experiment 2. The conjunction analysis shows that most probably the area actually was activated in Experiment 2, but the level of activation did not exceed the threshold for significance.

For the reading phase no significant differences between the two experiments were observed. The prediction that the left IFG and the dmPFC might show greater activation for reading under verification instruction than under recognition instructions was not confirmed. This prediction was put forward in a study by Chow et al. (2008) who varied reading instructions within subjects and thus employed tests with greater statistical power. Another difference to the study of Chow et al. is that their participants were instructed to either predict the upcoming events or to engage in "normal reading". It is possible that reading with the more focussed processing goal of recognition in Experiment 2 required greater contributions of the respective brain regions than normal reading without further processing goals.

The brain areas which were commonly activated during the execution of the verification and recognition tests—left IFG and left STS—formed a subset of the network that was activated for the conjunction of the reading phases of the experiments. The dmPFC was not found to be significantly activated in this analysis. However, it has to be noted that the event length, covered by the test regressor (1.8 seconds) was shorter than that of the reading regressor (3.6 seconds), and thus, statistical power was smaller.

The results of the direct contrast between verification and recognition task provide converging evidence for the central role of the dmPFC for situation level processing. The dmPFC and the pSTG/SMG were the only areas that were found to be greater activated during the verification task than during the recognition task. Presumably, in the verification task situation level processing is enhanced because participants need to evaluate the plausibility of the test statements in the given context. The results of the present studies as well as findings from other research, suggested that the contribution of the dmPFC can be interpreted as an indication of intentional, evaluative processing based on the utilization of the participants´ prior knowledge (Addis et al., 2007; Buckner & Carroll, 2007; Ferstl & von Cramon, 2002). Higher dmPFC activation in the verification experiment than in the recognition experiment is consistent with this interpretation.

The left temporo-parietal junction area was also found to be more activated during verification than during recognition. Functionally, this regions has been related to phonological processing in the context of verbal working memory and also to semantic processing. One interpretation is that the pSTG/SMG area provides initial access to semantic representations (“semantic activation”, Jung-Beeman, 2005) based on phonological representations (Hickok & Poeppel, 2004). The results of the direct comparison suggest these initial semantic processes were more intense during the verification test. This is compatible with the view that for verification accessing word meanings on the propositional level was more task relevant than for recognition.

E. General Discussion

The goal of the presented experiments was to refine proposals about the cortical network which implements inference processing in text comprehension. Starting points were considerations about (a) the cognitive representations of texts, (b) recognition and verification processes, and (c) existing hypotheses about the functional neuroanatomy of language comprehension processes. The main finding was that the dmPFC was more strongly related to inference processing than any other cortical region. The dmPFC was the only area which responded more to inference verification statements than to paraphrase statements in Experiment 1. Under recognition instructions in Experiment 2, the dmPFC also contributed to language processing, albeit to a lesser degree. This was also confirmed by direct comparisons between the verification and recognition tasks.

These results are in line with recent meta-analysis data on neuroimaging studies on language comprehension (Ferstl et al., 2007). Activation in the dmPFC was reliably found during coherence building processes in several studies. Considering a broader scope, it is interesting to note that the medial frontal cortex is one of the regions which are thought to be important for the mediation between past experience and future events. Recently, it has been emphasized by several authors (Buckner & Carroll, 2007; Schacter & Addis, 2007a) that an important function of episodic memory—if not the primary one—is the ability to simulate and predict future event. This has also been named *mental time travel* (Suddendorf & Corballis, 2007). Addis et al. (2007), for example, asked participants to generate and elaborate on either past events or imagined future events. During elaboration, both tasks activated medial aspects of the prefrontal and parietal cortex as well as areas in the left temporal lobe. Addis et al. conclude that the representation of past and future events is partially mediated by common neural substrates. With respect to the adaptive value of such a system, the authors argue that "Not only does the episodic system permit one to retrieve past episodes for evaluation regarding future approach or avoidance of similar scenarios, it also allows for the simulation of novel events in considerable detail [...]" (p.1374).

This development in the memory literature and the general notion of "simulations" as fundamental units of cognition (Barsalou, 1999) is also reflected in the immersed experiencer framework (IEF) in the field of language comprehension (see section A.1.3; Zwaan, 2004). The IEF posits that language provides the cues for the simulation of experientially based representations. The medial frontal areas which are reported to subserve the simulation of future events in the memory literature (Addis et al., 2007; Buckner & Carroll, 2007) are in close proximity to and partially overlapping with the dmPFC activations found in the present experiments. It is therefore tempting to conclude that the underlying mechanisms for mental time travel and the inference processes studied here might share a common basis. What exactly these common processes might be, is a similarly challenging question as how to precisely characterize mental simulations. As a common denominator, in both cases a certain scenario is evaluated against the background of prior experience, and the medial PFC might play a central role in these evaluation processes. This notion is compatible with the interpretation derived from Experiment 1—namely that the dmPFC was involved in the evaluation of the plausibility of the test statement in the given context. Note however, that despite some overlap most of the dmPFC activations found in Experiment 1 and Experiment 2 tended to be located in the more posterior and dorsal parts of the medial PFC, corresponding to BA 9, whereas the areas presumably subserving mental time travel seem to fall more into BA 10 (Addis et al., 2007; Buckner & Carroll, 2007).

A second important result of the presented experiments concerns the contribution of right hemispheric areas during the reading and test phases. The data presented here suggest that the right aTL was routinely activated during the reading of the sentences but it did not contribute to the processing of inference statements in a pronounced way. Comparatively weak evidence for an involvement of this region during the verification of inference statements was only found in the ROI analysis of Experiment 1. Here, inference statements elicited greater activation than explicit statements, but the difference between inferences and paraphrases was not significant. This is consistent with the observation that despite the prominent view of right hemisphere areas playing a particularly important role for inferencing, the evidence found for this claim is rather inconsistent (Ferstl et al., 2007). Most likely, the function of the bilateral aTL is generally related to the integration of semantic information on sentence and discourse level.

Modulation of activity in response to inference and paraphrase statements additionally occurred in the vIFG (BA 47) in both experiments, presumably indicating higher demands of semantic selection and control processes for inferences than for paraphrases. Thus, within the network of brain regions that were commonly activated during the reading phases of both areas, inference processing was primarily associated with activation of the dmPFC and, to a smaller degree, with the vIFG (BA 47).

Figure 25 illustrates the language comprehension network sketched in this and previous chapters. The regions which were commonly activated during reading in both experiments are shown along with speculations about their possible functional roles. In summary, the TPJ is assumed to be important for the initial step in semantic processing—semantic activation (Jung-Beeman, 2005). This area has been described as mapping sound-based representations to meaning-based representations in temporal areas (Hickok & Poeppel, 2004), such as the mTL, which was also found in the present studies. The aTL seems to be involved in semantic integration processes, such as “propositionalization” (Ferstl & von Cramon, 2001) or “construal” (Zwaan, 2004). The vIFG is assumed to participate in controlling semantic retrieval processes (semantic selection; Wagner et al., 2001), whereas the dmPFC presumably contributes to comprehension by evaluation processes on the situational level. Activity in the PCC was associated to memory retrieval processes, and the vmPFC was related to emotion processing (Ferstl et al., 2007). Although right hemisphere areas are not shown in Figure 25, it is assumed that the specified regions of the left hemisphere are supported by their homologue right hemisphere areas. The precise nature of this support is still a matter of question. The current studies are compatible with the view that activations in right hemisphere regions primarily reflect increasing processing demands.

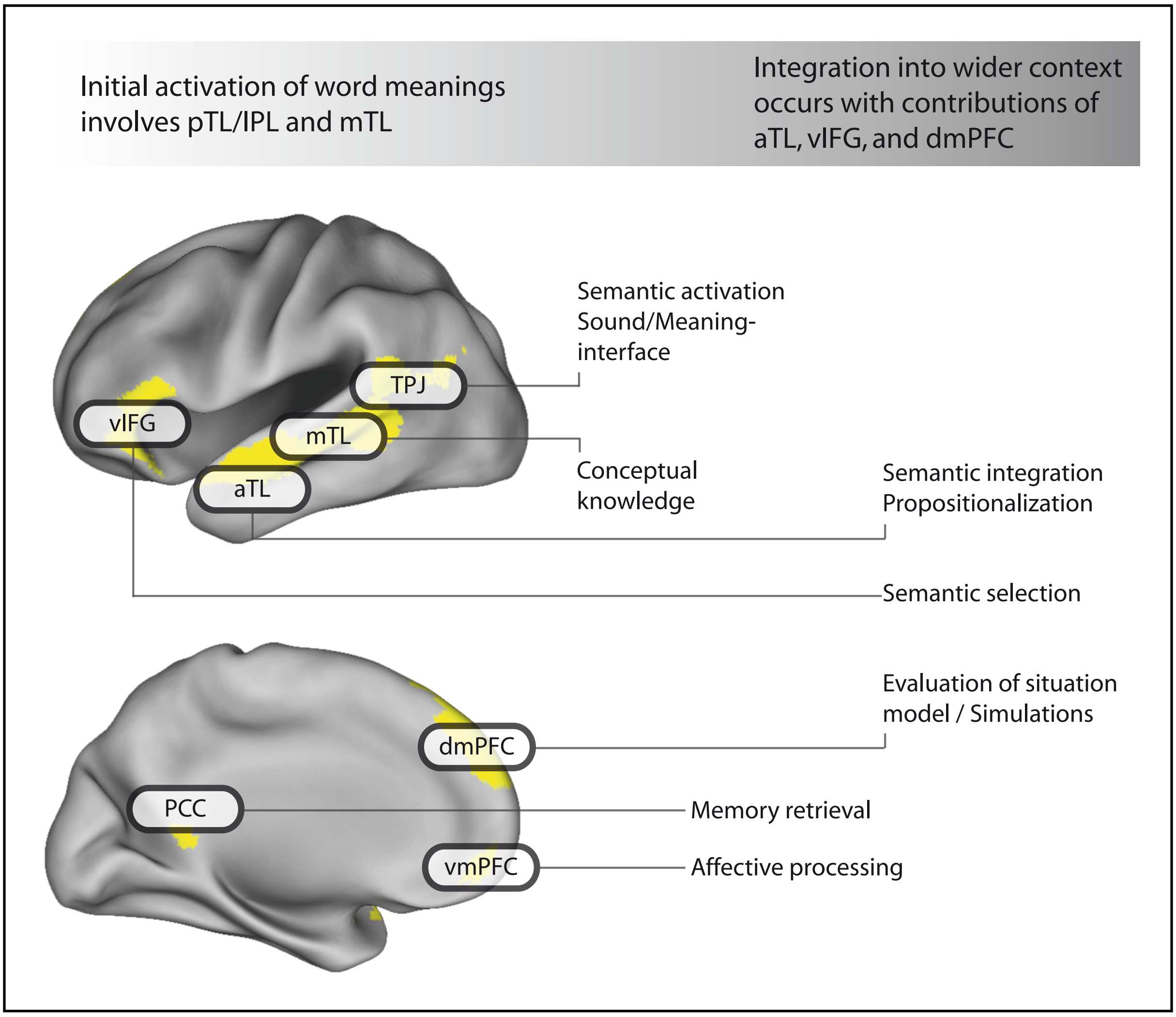


Figure 25: Proposed network for inference processing in the present experiments. Regions were selected on the basis of the results of the conjunction analysis of the reading phases of both experiments. Only left hemisphere areas are shown. Bilateral activations were found in the dmPFC, vmPFC, and aTL.

Thus, during reading or listening to language utterances the TPJ is assumed to be involved in activating the individual word meanings with contributions of the mTL. In case of ambiguities the vIFG might be the place where an interpretation is selected for further integration processes in the aTL. The PCC contributes to further memory retrieval processes. If affective components play a role for comprehension, the vmPFC is engaged. Finally, the dmPFC is presumably involved in the evaluation of the plausibility of the evolving situation model in the current context. To which extent inferences are actually drawn will depend on, amongst other things, the specific goals of the reader.

Note that this conceptualization is considerably different from the proposal of Mason and Just (2004). These authors suggested that a LH-language network supports basic language processed (lexical access, syntactical parsing, etc.) independent of coherence. They assumed that the dlPFC supports the generation of inferences, and that the integration of inferences occurs in a RH-language network. The results of the present experiments can not be explained satisfactorily by this model. First of all, modulations of activity due to the sentence condition manipulation was established in all lateral left hemisphere regions of interests in Experiment 1 and in most equivalent regions in Experiment 2. Secondly, activation in the dlPFC was not observed at all in Experiment 1. In Experiment 2 left dlPFC activity was found in the comparisons of language conditions versus the pseudoword condition as a baseline and most likely reflected differences in working memory demands between the tasks. Thirdly, right hemisphere activation in Experiment 1 was only found in concert with activation in the corresponding left hemisphere regions where it was generally more pronounced.

Of course, several words of caution also need to be noted concerning the presented experiments. Some important aspects are shortly discussed in the remainder of this thesis. First, the theoretical foundations of inference processes in general are relatively weak. Second, we still know comparatively little about the operating principles of the brain regions which enable language functions. In particular, the precise role of the right hemisphere has not been sufficiently clarified yet. Third, the used sentence materials left some room for improvement.

Concerning inference theory Graesser et al. (1994) noted that

> It is very tedious and difficult to specify the numerous knowledge structures associated with text, so theorists and empirical researchers have normally avoided a systematic analysis of world knowledge. Nevertheless, a mature theory of inference generation would need to analyze world knowledge in detail" (p. 376).

As has been put down by Frank, Koppen, Noordman, and Vonk (2003), most theories of text comprehension, like the CI-model, do not provide convincing mechanisms of inference construction processes because they do not explain how world knowledge is accessed. The authors criticized that, for instance, in implementations of the CI-model (like in Schmalhofer et al., 2002) the modelers selected the propositions which constitute plausible inferences in advance, and that the selection criteria were not well-defined. Frank et al. proposed a model—the *distributed situation space model* (DSS)—which overcomes this problem by representing text events as temporal sequences of vectors in a multi-dimensional "situation-state space" (Frank et al., 2003; Frank, Koppen, Noordman, & Vonk, 2007). In this model, a story can be viewed as a trajectory through the situation-state space, and for each point in that space, probability values can be calculated. Therefore, inferencing is reflected in the increased likelihood of certain situation vectors for a given state of the knowledge structure. A major drawback of this model is that, so far, it works only for a very limited knowledge base (the so-called microworld consisting of 14 basic propositions). Clearly, the DSS model has to demonstrate how it handles more plausible quantities of knowledge but the explicit attempt to explain inference processes on the basis of general knowledge structures is certainly promising.

Another important theoretical development is the notion of simulations in the immersed experiencer framework (IEF, Zwaan, 2004). This approach emphasizes the role of prior experience for language comprehension and provides a bridge to recent interpretations about the role of episodic memory in mental simulations of the future. Furthermore, some of its key concepts are in part motivated by neuroscientific evidence (perceptual symbols, Barsalou, 1999; functional webs, Pulvermuller, 1999). Unfortunately, apart from providing an interesting and inspiring perspective on language comprehension and episodic memory, the notion of simulation processes as such is similarly vague as the notion that inferences arise from general knowledge structures (Van Dijk & Kintsch, 1983). Although the IEF identifies and explicitly includes several important factors which can influence comprehension processes (e.g., spatial, temporal, and perspective shifts in the situation model), further specifications of the framework are necessary. With respect to inferences processes it remains unclear which aspects of a situation are actually included in a simulation during the construal process. In other words, the question how inferences—as parts of our general knowledge—are activated and selected for the interpretation of an utterance still remains largely unanswered.

Given that most, if not all, cognitive functions play a role for language comprehension, designing neuroimaging experiments in this field is a tedious and complex task, calling for highly interdisciplinary efforts from psychologists, linguists, and neuroscientists. The risk of confounding the experimental manipulations of interest with unwanted factors is relatively high because the language materials usually underly plenty of restrictions. In the current studies one experimental confound had been identified and discussed in the context of Experiment 2, namely that paraphrase sentences contained less common wordings than explicit sentences. It is not difficult to imagine that the sentence materials contained other potentially confounding characteristics. For instance, inference sentences had a different syntactic structure than explicit and paraphrase sentences because instead of mentioning the explicit outcome of a situation, in most cases elaborations of the situation were stated in additional subordinate clauses. Whether participants used this information strategically, e.g. to predict the upcoming requested yes- or no-response, can not be ruled out completely. Although it is reasonable to argue that response times should not differ if responses were predictable, other interpretations might be possible.

Experimental manipulation in language comprehension studies are likely to engage brain networks which mediate several cognitive functions. This makes neuroimaging studies particularly vulnerable for experimental confounds and misinterpretations. Without (a) well-formulated theories which allow specifications of the underlying cognitive processes, (b) carefully controlled materials, and (c) extensive knowledge about the brain areas and mechanisms which support the cognitive processes under question, interpreting neuroimaging results depends to considerable degree on the researchers' and readers' willingness to accept a certain amount of speculations. A lot of truly interdisciplinary research will still be necessary to provide deeper understanding of the neural basis of language comprehension processes.

F. Conclusion

This research has provided further evidence for the central role of the dmPFC for inference processes. Besides relating the results to the neuroimaging literature in the language comprehension domain, the attempt was made to additionally incorporate findings from research on more general memory processes. Particularly interesting is the convergence of findings that medial portions of the prefrontal cortex seem to be important for both mental simulations and inference processes. This finding is relevant for theories on language comprehension because it provides further evidence for approaches like the immersed experiencer framework (Zwaan, 2004) which emphasize the role of mental simulations for language comprehension instead of the manipulation of amodal representations.

G. References

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*(7), 1363-77.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, *22*(4), 577-660.

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain and Language*, *44*(1), 80-120.

Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In M. Beeman, & C. Chiarello (Eds.), *Right hemisphere language comprehension : perspectives from cognitive neuroscience*. (pp. 255-84). Mahwah, N.J.: L. Erlbaum Associates.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151-88.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P. et al. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language. A positron emission tomography activation study. *Brain*, *117(*6), 1241-53.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*, 193-209.

Bransford, J. D. & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, *2*(4), 331-50.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). *Region of interest analysis using an SPM toolbox*. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Available on CD-ROM in NeuroImage, 16(2), 497.

Buckner, R. L. & Wheeler, M. E. (2001). The cognitive neuroscience of remembering. *Nature Reviews Neuroscience*, *2*(9), 624-34.

Buckner, R. L. (2003). Functional-anatomic correlates of control processes in memory. *Journal of Neuroscience*, *23*(10), 3999-4004.

Buckner, R. L. & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Science*, *11*(2), 49-57.

Cabeza, R. & Nyberg, L. (2000a). Neural bases of learning and memory: functional neuroimaging evidence. *Current Opinion in Neurology*, *13*(4), 415-21.

Cabeza, R. & Nyberg, L. (2000b). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*(1), 1-47.

Calvo, M. G., Castillo, M. D., & Schmalhofer, F. (2006). Strategic influence on the time course of predictive inferences in reading. *Memory & Cognition*, *34*(1), 68-77.

Carpenter, P. A. & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45-73.

Carreiras, M., Mechelli, A., Estevez, A., & Price, C. J. (2007). Brain activation for lexical decision and reading aloud: two sides of the same coin? *Journal of Cognitive Neuroscience*, *19*(3), 433-44.

Chow, H. M., Kaup, B., Raabe, M., & Greenlee, M. W. (2008). Evidence of fronto-temporal interactions for strategic inference processes during language comprehension. *Neuroimage*, *40*(2), 940-54.

Cohen, M. S. (1996). Rapid MRI and Functional Applications. In A. W. Toga, & J. C. Mazziotta (Eds.), *Brain Mapping: The Methods*. San Diego: Academic Press.

Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? *Cortex*, *42*(3), 323-31.

Connell, L. & Keane, M. T. (2006). A Model of Plausibility. *Cognitive Science*, *Vol 30*(1), 95-120.

Cook, A. E., Limber, J. E., & Edward, O. B. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language*, *44*(2), 220-30.

Corbett, A. T. & Dosher, B. A. (1978). Instrument inferences in sentence encoding. *Journal*

of Verbal Learning & Verbal Behavior, *17*(4), 479-81.
Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, *126*(Pt 5), 1193-201.
D'Arcy, R. C., Ryner, L., Richter, W., Service, E., & Connolly, J. F. (2004). The fan effect in fMRI: left hemisphere specialization in verbal working memory. *Neuroreport*, *15*(12), 1851-55.
Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition: International Journal of Cognitive Science*, *92*(1-2), 179-229.
Ferstl, E. C. (2007a). The functional neuroanatomy of text comprehension: What's the story so far? In F. Schmalhofer, & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. Mahwah, NJ: Erlbaum.
Ferstl, E. C., Rinck, M., & von Cramon, D. Y. (2005). Emotional and Temporal Aspects of Situation Model Processing during Text Comprehension: An Event-Related fMRI Study. *Journal of Cognitive Neuroscience*, *17*(5), 724-39.
Ferstl, E. C. & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *Neuroimage*, *17*(3), 1599-612.
Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, Y. D. (2007). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581-93.
Ferstl, E. C. (2007b). The functional neuroanatomy of text comprehension: What's the story so far. In F. Schmalhofer, & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. (pp. 53-102). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
Ferstl, E. C. & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research*, *11*, 325-40.
Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*(4), 981-95.
Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes*, *22*(3), 225-46.
Fletcher, C. R. & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. *Discourse Processes*, *13*(2), 175-90.
Fletcher, P. C., Happe, F., Frith, U., & Baker, S. C. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition: International Journal of Cognitive Science*, *57*(2), 109-28.
Fodor, J. (1999). Let your brain alone. *LRB* Retrieved March, 2008, from http://www.lrb.co.uk/v21/n19/fodo01_.html
Forstmann, B. U., Brass, M., Koch, I., & von Cramon, D. Y. (2005). Internally generated and directly cued task sets: an investigation with fMRI. *Neuropsychologia*, *43*(6), 943-52.
Frank, S. L., Koppen, M., Noordman, L. G., & Vonk, W. (2007). Coherence-driven resolution of referential ambiguity: a computational model. *Memory & Cognition*, *35*(6), 1307-22.
Frank, S. L., Koppen, M., Noordman, L. G. M., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science: A Multidisciplinary Journal*, *27*(6), 875-910.
Friese, U., Rutschmann, R. M., Raabe, M., & Schmalhofer, F. (in press). Neural Indicators of Inference Processes in Text Comprehension: An Event-Related fMRI-Study. *Journal of Cognitive Neuroscience*.
Friston, K. J. (1996). Statistical Parapetric Mapping and Other Analyses of Functional Imaging Data. In A. W. Toga, & J. C. Mazziotta (Eds.), *Brain Mapping: The Methods*. San Diego: Academic Press.

Friston, K. J., Penny, W. D., & Glaser, D. E. (2005). Conjunction revisited. *Neuroimage*, *25*(3), 661-67.
Frith, C. D. & Frith, U. (1999). Interacting minds--a biological basis. *Science*, *286*, 1692-95.
Gallagher, H. L. & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*(2), 77-83.
Gernsbacher, M. A. & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual Review of Psychology*, *54*, 91-114.
Gilboa, A. (2004). Autobiographical and episodic memory--one and the same? Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia*, *42*(10), 1336-49.
Gillund, G. & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1-67.
Gitelman, D. R., Nobre, A. C., Sonty, S., Parrish, T. B., & Mesulam, M. M. (2005). Language network specializations: an analysis with parallel task designs and functional magnetic resonance imaging. *Neuroimage*, *26*(4), 975-85.
Glenberg, A. M. & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, *43*(3), 379-401.
Goel, V., Gold, B., Kapur, S., & Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *Neuroreport*, *8*(5), 1305-10.
Gordon, A. M., Lee, J. H., Flament, D., Ugurbil, K., & Ebner, T. J. (1998). Functional magnetic resonance imaging of motor, sensory, and posterior parietal cortical areas during performance of sequential typing movements. *Exp Brain Res*, *121*(2), 153-66.
Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163-89.
Graesser, A. C., Louwerse, M. M., McNamara, D. S., Olney, A., Cai, Z., & Mitchell, H. H. (2007). Inference generation and cohesion in the construction of situation models: Some connections with computational linguistics. In F. Schmalhofer, & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. (pp. 289-310). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371-95.
Graham, K. S., Lee, A. C., Brett, M., & Patterson, K. (2003). The neural basis of autobiographical and semantic memory: new evidence from three PET studies. *Cognitive, Affective & Behavioral Neuroscience*, *3*(3), 234-54.
Griesel, C., Friese, U., & Schmalhofer, F. (2003). What are the Differences in the Cognitive Representations of Predictive and Bridging Inferences? In F. Schmalhofer, R. Young, & G. Katz (Eds.), *Proceedings of EuroCogSci03: The European Cognitive Science Conference 2003*. (pp. 145–50). Mahwah, New Jersey: Lawrence Erlbaum Associates.
Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416-23.
Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, *21*(4), 1639-51.
Haviland, S. E. & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning & Verbal Behavior*, *13*(5), 512-21.
Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Expimental Psychology*, *58*(2), 193-233.
Henson, R. (2006). What has (neuro)psychology told us about the mind (so far)? A reply to Coltheart (2006). *Cortex*, *42*(3), 387-92.
Henson, R. N., Hornberger, M., & Rugg, M. D. (2005). Further dissociating the processes involved in recognition memory: an FMRI study. *Journal of Cognitive Neuroscience*, *17*(7), 1058-73.
Hernandez, L., Wager, T. D., & Jonides, J. (2002). Introduction to Functional Brain Imaging.

In J. Wixted, & H. Pashler (Eds.), *Stevens Handbook of Experimental Psychology: Third Edition, Volume 4: Methodology in Experimental Psychology*. New York: John Wiley and Sons,Inc.

Hickok, G. & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition: International Journal of Cognitive Science*, *92*(1-2), 67-99.

Indefrey, P. & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition: International Journal of Cognitive Science*, *92*(1-2), 101-44.

Jonides, J., Nee, D. E., & Berman, M. G. (2006). What has functional neuroimaging told us about the mind? So many examples, so little space. *Cortex*, *42*(3), 414-7; discussion 422-7.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Science*, *9*(11), 512-18.

Jung-Beeman, M., Bowden, E. M., & Gernbacher, M. A. (2000). Right and left hemisphere cooperation for drawing predictive and coherence interferences during normal story comprehension. *Brain & Language*, *71*(2), 310-36.

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. *Journal of Neuroscience*, *24*(17), 4172-80.

Keefe, D. E. & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory & Language*, *32*, 446-63.

Kiefer, M., Weisbrod, M., Kern, I., Maier, S., & Spitzer, M. (1998). Right hemisphere activation during indirect semantic priming: Evidence from event-related potentials. *Brain & Language*, *64*(3), 377-408.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163-82.

Kintsch, W. (1998). *Comprehension : a paradigm for cognition*. Cambridge, U.K. ; New York: Cambridge University Press.

Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363-94.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*(2), 133-59.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience*, *15*(2), 272-93.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of Discourse: an fMRI study of causal inferencing across sentences. *Neuroimage*, *33*, 343-61.

Lea, R. B., Mulligan, E. J., & Walton, J. L. (2005). Accessing Distant Premise Information: How Memory Feeds Reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *Vol 31*(3), 387-95.

Li, C. S., Huang, C., Constable, R. T., & Sinha, R. (2006). Imaging response inhibition in a stop-signal task: neural correlates independent of signal monitoring and post-response processing. *Journal of Neuroscience*, *26*(1), 186-92.

Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *Journal of Neuroscience*, *23*(10), 3963-71.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Magliano, J. P., Radvansky, G. A., & Copeland, D. E. (2007). Beyond Language Comprehension: Situation models as a form of autobiographical memory. In F. Schmalhofer, & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. Mawah, NJ: Lawrence Erlbaum Associates.

Mar, R. A. (2004). The neuropsychology of narrative: story comprehension, story production and their interrelation. *Neuropsychologia*, *42*(10), 1414-34.

Mason, R. A. & Just, M. A. (2004). How the brain processes causal inferences in text. *Psychological Science*, *15*(1), 1-7.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O. et al. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*(4), 467-79.

McDaniel, M. A., Schmalhofer, F., & Keefe, D. E. (2001). What is minimal about predictive inferences? *Psychonomic Bulletin & Review*, *8*(4), 840-46.

McDermott, K. B., Jones, T. C., Petersen, S. E., Lageman, S. K., & Roediger, H. L., 3rd. (2000). Retrieval success is accompanied by enhanced activation in anterior prefrontal cortex during recognition memory: an event-related fMRI study. *Journal of Cognitive Neuroscience*, *12*(6), 965-76.

McKoon, G. & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *12*, 82-91.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440-66.

Mechelli, A., Crinion, J. T., Long, S., Friston, K. J., Lambon Ralph, M. A., Patterson, K. et al. (2005). Dissociating reading processes on the basis of neuronal interactions. *Journal of Cognitive Neuroscience*, *17*(11), 1753-65.

Myers, J. L. & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *Vol 26*(2-3), 131-57.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences USA*, *87*(24), 9868-72.

Ollinger, J. M., Corbetta, M., & Shulman, G. L. (2001). Separating processes within a trial in event-related functional MRI. *Neuroimage*, *13*(1), 218-29.

Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001). Separating processes within a trial in event-related functional MRI. *Neuroimage*, *13*(1), 210-17.

Page, M. P. (2006). What can't functional neuroimaging tell the cognitive psychologist? *Cortex*, *42*(3), 428-43.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci*, *10*(2), 59-63.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, *10*(1), 15-35.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory & Language*, *27*(4), 399-415.

Pulvermuller, F. (1999). Words in the brain's language. *Behavioral & Brain Sciences*, *22*(2), 253-336.

Pulvermuller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, *5*(12), 517-24.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc Natl Acad Sci U S A*, *98*(2), 676-82.

Raichle, M. E. & Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage*, *37*(4), 1083-90; discussion 1097-9.

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, *Vol 89*(3), 250-80.

Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: Common and distinct neuronal processes. *Neuropsychologia*, *41*(9), 1241-53.

Savoy, R. L. (2005). Experimental design in brain activation MRI: cautionary tales. *Brain Research Bulletin*, *67*(5), 361-67.

Schacter, D. L. & Addis, D. R. (2007a). Constructive memory: the ghosts of past and future.

Nature, *445*(7123), 27.
Schacter, D. L. & Addis, D. R. (2007b). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Societyof London B: Biological Sciences*, *362*(1481), 773-86.
Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, *33*(2), 105-32.
Schmalhofer, F. & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory & Language*, *25*(3), 279-94.
Schnyer, D. M., Nicholls, L., & Verfaellie, M. (2005). The role of VMPC in metamemorial judgments of content retrievability. *Journal of Cognitive Neuroscience*, *17*(5), 832-46.
Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, *54*(4), 574-91.
Singer, M. & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning & Verbal Behavior*, *22*(4), 437-48.
Singer, M., & Leon, J. (2007). Psychological studies of higher language processes: Behavioral and empirical approaches. In F. Schmalhofer, & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. (pp. 9-25). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34-50.
St George, M., Kutas, M., Martinez, A., & Sereno, M. I. (1999). Semantic integration in reading: Engagement of the right hemisphere during discourse processing. *Brain*, *122*(7), 1317-25.
Stowe, L. A., Haverkort, M., & Zwarts, F. (2004). Rethinking the neurological basis of language. *Lingua: International Review of General Linguistics*, *115*(7), 997-1042.
Suddendorf, T. & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*(3), 299-313.
Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia*, *44*(12), 2189-208.
Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences USA*, *94*(26), 14792-97.
Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/ procedural model. *Cognition: International Journal of Cognitive Science*, *92*(1-2), 231-70.
Van den Broek, P., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A 'landscape' view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton, & A. C. Graesser (Eds.), *Models of understanding text*. (pp. 165-87). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
Van den Broek, P., Rohleder, L., & Narvaez, D. (1994). Cognitive processes in the comprehension of literary texts. In H. Van Oostendorp, & R. A. Zwaan (Eds.), *Naturalistic text comprehension*. (pp. 229-46). Noewood, New Jersey: Ablex Publishing.
Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. San Diego, CA: Academic Press.
Van Essen, D. C. (2002). Windows on the brain: the emerging role of atlases and databases in neuroscience. *Current Opinion in Neurobiology*, *12*(5), 574-79.
Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of American Medical Informatics Association*, *8*(5), 443-59.
Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to

sentences. *Journal of Cognitive Neuroscience*, *14*(4), 550-60.
Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Jung Beeman, M. (2006). Neural activity of inferences during story comprehension. *Brain Research*, *1084*(1), 104-14.
Wager, T. D., Hernandez, L., Jonides, J., & Lindquist, M. (2007). Elements of functional neuroimaging. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology, 4th Ed*. (pp. 19-55). Cambridge: Cambridge University Press.
Wager, T. D. (2006). Do We Need to Study the Brain to Understand the Mind? *APS Observer*, *19*(9), 25-27.
Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*(2), 329-38.
Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445-53.
Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within 'Wernicke's area'. *Brain*, *124*(Pt 1), 83-95.
Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage*, *25*(3), 1002-15.
Zwaan, R. A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation, 44*. (pp. 35-62). New York: Academic Press.
Zwaan, R. A., & Singer, M. (2003). Text comprehension. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes*. (83-121). Mawah, New Jersey: Lawrence Erlbaum Associates.
Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-85.

H. Appendices

H.1. Instructions Experiment 1

The following instructions were presented to the participants in the training period outside the scanner room. A shorter version was presented additionally inside the scanner.

Vielen Dank für Ihre Teilnahme an dieser Studie!

Bitte lesen Sie sich diese Instruktionen genau durch. Wenn Sie Fragen haben, wenden Sie sich an die Versuchsleiterin oder den Versuchsleiter.

In dieser Untersuchung geht es um die Frage, wie das Gehirn Texte verarbeitet. Dieses Experiment ist in drei Blöcke unterteilt und jeder Block besteht dabei aus 36 Aufgaben. Eine Aufgabe sieht so aus:

Sie lesen zunächst einen kurzen Text, der Ihnen Wort für Wort im Zentrum des Bildschirms präsentiert wird. Jeder Text wird mit einer Überschrift aus zwei Worten eingeleitet. Die Überschrift beschreibt den Kontext bzw. das Thema des folgenden Textes. Damit Sie das Ende des Textes erkennen können, befindet sich hinter dem letzen Wort ein "*". Unmittelbar nach dem Text erscheinen auf dem Bildschirm zwei Worte, die eine bestimmte Aussage bilden, z.B. "Essen verdorben". Sie sollen dann überlegen, ob diese Aussage in Bezug auf die zuvor beschriebene Situation zutrifft oder nicht zutrifft.

Sobald Sie die Worte der Aussage gelesen haben, geben Sie bitte möglichst zügig Ihre Antwort mithilfe der Reaktionstasten. Legen Sie bitte Zeigefinger und Mittelfinger ihrer dominanten Hand über die beiden Antworttasten, so dass Sie bequem und einfach die gewünschte Taste drücken können.

Drücken Sie die linke Taste, wenn die Aussage zutrifft und drücken Sie die rechte Taste, wenn die Aussage nicht zutrifft. Wenn Sie reagiert haben, verschwindet das Wortpaar und unmittelbar darauf erscheint ein Kreuz in der Mitte des Bildschirmes, welches Sie bitte betrachten. Nach einer kurzen Zeit wird der nächste Satz präsentiert.

Achtung: In einigen Durchgängen werden Ihnen Folgen sinnloser Wörter dargeboten. Es ist sehr wichtig, dass Sie diese "Pseudo-Wort"-Sätze genauso lesen als wären es echte Worte! Ihre Aufgabe bei den "Pseudo-Wort"-Sätzen ist es anzugeben, ob die beiden Pseudo-Worte, die Ihnen als Wortpaar gezeigt werden, dieselben sind wie die beiden letzten des zuvor präsentierten Satzes. Die Antworten der Pseudo-Wort Sätze beziehen Sie sich also nur auf die beiden zuletzt präsentierten Worte des Satzes. Drücken Sie die linke Taste, wenn das "Pseudo-Wortpaar" den letzen beiden Worten des Satzes entsprach und drücken Sie die rechte Taste, wenn dies nicht der Fall war.

In den Zeiten, in denen es nichts zu lesen gibt, erscheint im Zentrum des Bildschirms ein Kreuz. Da Augenbewegungen bei der Aufzeichnung der Hirnaktivität Störungen erzeugen können, müssen Sie bitte unbedingt zwischen den Textpräsentationen auf diese Zeichen im Zentrum blicken!

Die gesamte Untersuchung ist in drei Blöcke aufgeteilt und wird etwa eine Stunde dauern. Damit Sie sich zwischendurch etwas ausruhen können, gibt es nach jedem Block eine kurze Pause. In diesen Pausen können Sie die Augen bewegen oder schließen, müssen jedoch weiterhin ruhig liegen bleiben.

Sie erhalten nach jedem Block auch einen Bericht darüber, wie viele Ihrer Antworten richtig und wie viele falsch waren. Versuchen Sie bitte auf jeden Fall, fehlerfrei und so schnell wie möglich zu reagieren, da wir Datensätze mit zu vielen falschen oder zu langsamen Antworten leider nicht auswerten können!

Hier ist noch einmal der genaue Ablauf eines Durchgangs in Kurzform:

1. Wort für Wort Präsentation eines Textes.
2. Präsentation der Worte der Aussage
3. Aufzeichnung ihrer Antwort
4. Kurze Pause mit Kreuz in der Mitte des Bildschirms.
5. Der nächste Satz wird gezeigt
6. usw.

Wenn Sie keine Fragen haben, beginnen wir zunächst mit ein paar Probedurchgängen außerhalb des Kernspintomografen. Nach den Probedurchgängen können auch noch einmal Fragen an den Versuchsleiter gestellt werden.

Zur Information:

Nach dem Experiment wird zusätzlich eine hochauflösende Aufnahme ihres Gehirnes durchgeführt. Das dauert ungefähr zehn Minuten. Währenddessen können sie ihre Augen bewegen oder schließen, müssen jedoch weiterhin ruhig liegen bleiben.

Zum Starten der Übung betätigen Sie eine der beiden Reaktionstasten!

H.2. Instruction Experiment 2

The following instructions were presented to the participants in the training period outside the scanner room. A shorter version was presented additionally inside the scanner.

Vielen Dank für Ihre Teilnahme an dieser Studie!

Bitte lesen Sie sich diese Instruktionen genau durch. Wenn Sie Fragen haben, wenden Sie sich an die Versuchsleiterin oder den Versuchsleiter.

In dieser Untersuchung geht es um die Frage, wie das Gehirn Texte verarbeitet. Dieses Experiment ist in drei Blöcke unterteilt und jeder Block besteht dabei aus 36 Aufgaben. Eine Aufgabe sieht so aus:

Sie lesen zunächst einen kurzen Text, der Ihnen Wort für Wort im Zentrum des Bildschirms präsentiert wird. Jeder Text wird mit einer Überschrift aus zwei Worten eingeleitet. Die Überschrift beschreibt den Kontext bzw. das Thema des folgenden Textes. Damit Sie das Ende des Textes erkennen können, befindet sich hinter dem letzen Wort ein "*". Unmittelbar nach dem Satz erscheinen auf dem Bildschirm zwei oder drei Worte, die eine bestimmte Aussage bilden, z.B. "Essen verdorben".

Sie sollen dann überlegen, ob alle Worte im zuvor gelesenen Text erwähnt wurden oder nicht. Ein Wort gilt als erwähnt, sobald der Wortstamm im Text vorkam. Beispiele für Aussagen, deren Wortpaare als im Text erwähnt gelten:

Beispiel 1)
Die Email.
Pascal schrieb Sonja eine Email.*
Aussage: "Email geschrieben

Beispiel 2)
Das Fußballspiel.
Mirko lobte die Mannschaft nach dem Spiel.*
Aussage: "Mannschaft gelobt"

Sobald Sie die Worte der Aussage gelesen haben, geben Sie bitte möglichst zügig Ihre Antwort mithilfe der Reaktionstasten. Legen Sie bitte Zeigefinger und Mittelfinger ihrer dominanten Hand über die beiden Antworttasten, so dass Sie bequem und einfach die gewünschte Taste drücken können.

Drücken Sie die linke Taste, wenn alle Worte der Aussage im Text erwähnt wurden und drücken Sie bitte die rechte Taste, wenn nicht alle Worte der Aussage in dem zuvor gelesenen Text vorkamen. Wenn Sie reagiert haben, verschwindet das Wortpaar und unmittelbar darauf erscheint ein Kreuz in der Mitte des Bildschirmes, welches Sie bitte betrachten. Nach einer kurzen Zeit wird der nächste Satz präsentiert.

Achtung: In einigen Durchgängen werden Ihnen Folgen sinnloser Wörter dargeboten. Es ist sehr wichtig, dass Sie diese "Pseudo-Wort"-Sätze genauso lesen als wären es echte Worte! Ihre Aufgabe bei den "Pseudo-Wort"-Sätzen ist es anzugeben, ob die beiden Pseudo-Worte, die Ihnen als Wortpaar gezeigt werden, dieselben sind wie die beiden letzten des zuvor präsentierten Satzes. Die Antworten der Pseudo-Wort Sätze beziehen Sie sich also nur auf die beiden zuletzt präsentierten Worte des Satzes. Drücken Sie die linke Taste, wenn das "Pseudo-Wortpaar" den letzen beiden Worten des Satzes entsprach und drücken Sie die rechte Taste, wenn dies nicht der Fall war.

In den Zeiten, in denen es nichts zu lesen gibt, erscheint im Zentrum des Bildschirms ein Kreuz. Da Augenbewegungen bei der Aufzeichnung der Hirnaktivität Störungen erzeugen können, müssen Sie bitte unbedingt zwischen den Textpräsentationen auf diese Zeichen im Zentrum blicken!

Die gesamte Untersuchung ist in drei Blöcke aufgeteilt und wird etwa eine Stunde dauern. Damit Sie sich zwischendurch etwas ausruhen können, gibt es nach jedem Block eine kurze Pause. In diesen Pausen können Sie die Augen bewegen oder schließen, müssen jedoch weiterhin ruhig liegen bleiben.

Sie erhalten nach jedem Block auch einen Bericht darüber, wie viele Ihrer Antworten richtig und wie viele falsch waren. Versuchen Sie bitte auf jeden Fall, fehlerfrei und so schnell wie möglich zu reagieren, da wir Datensätze mit zu vielen falschen oder zu langsamen Antworten leider nicht auswerten können!

Hier ist noch einmal der genaue Ablauf eines Durchgangs in Kurzform:

1. Wort für Wort Präsentation eines Textes.
2. Präsentation der Worte der Aussage
3. Aufzeichnung ihrer Antwort
4. Kurze Pause mit Kreuz in der Mitte des Bildschirms.
5. Der nächste Satz wird gezeigt
6. usw.

Wenn Sie keine Fragen haben, beginnen wir zunächst mit ein paar Probedurchgängen außerhalb des Kernspintomografen. Nach den Probedurchgängen können auch noch einmal Fragen an den Versuchsleiter gestellt werden.

Zur Information:

Nach dem Experiment wird zusätzlich eine hochauflösende Aufnahme ihres Gehirnes durchgeführt. Das dauert ungefähr zehn Minuten. Währenddessen können sie ihre Augen bewegen oder schließen, müssen jedoch weiterhin ruhig liegen bleiben.

Zum Starten der Übung betätigen Sie eine der beiden Reaktionstasten!

H.3. Experimental text materials

Abbreviations: EX - explicit, PA - paraphrase, IN - inference, UN - unrelated

Headline	Context sentence variations		Test statement
Am Strand	EX	Als sich Sabine bei den Felsen aufhielt, trat sie auf eine Glasscherbe und schnitt sich in den Fuß.	In Fuß geschnitten
Am Strand	PA	Als sich Sabine bei den Felsen aufhielt, trat sie auf eine Glasscherbe und ritzte sich den Fuß auf.	In Fuß geschnitten
Am Strand	IN	Als sich Sabine bei den Felsen aufhielt, trat sie auf eine Glasscherbe, die scharfe Kanten und Spitzen hatte.	In Fuß geschnitten
Am Strand	UN	Als sich Sabine bei den Felsen aufhielt, fand sie eine schöne Glasscherbe, die eine seltsame Aufschrift eingraviert hatte	In Fuß geschnitten
Das alte Haus	EX	Das Haus war alt und marode. Es konnte dem Erdbeben nicht stand halten und stürzte schließlich vollständig ein.	Haus eingestürzt
Das alte Haus	PA	Das Haus war alt und marode. Es konnte dem Erdbeben nicht stand halten und brach in sich zusammen.	Haus eingestürzt
Das alte Haus	IN	Das Haus war alt und marode. Es konnte dem Erdbeben nicht stand halten, welches zudem ungewöhnlich stark war.	Haus eingestürzt
Das alte Haus	UN	Das Haus war alt und marode. Es konnte dem Besitzer nichts anderes übrig bleiben, als es zu renovieren.	Haus eingestürzt
Das Bild	EX	Janine hatte auf dem Flohmarkt ein Bild gekauft. Sie nahm Hammer und Nagel und hängte es gleich auf.	Bild aufgehängt
Das Bild	PA	Janine hatte auf dem Flohmarkt ein Bild gekauft. Sie nahm Hammer und Nagel und brachte es gleich an.	Bild aufgehängt
Das Bild	IN	Janine hatte auf dem Flohmarkt ein Bild gekauft. Sie nahm Hammer und Nagel und wählte eine passende Stelle.	Bild aufgehängt
Das Bild	UN	Janine hatte auf dem Flohmarkt ein Bild gekauft. Sie hatte vor, es ihrem Vater zum Geburtstag zu schenken.	Bild aufgehängt
Das Unwetter	EX	Als an einem Abend der Strom ausfiel, holte Nina aus der Schublade eine Kerze und zündete sie an.	Kerze angezündet
Das Unwetter	PA	Als an einem Abend der Strom ausfiel, holte Nina aus der Schublade eine Kerze und steckte sie an.	Kerze angezündet
Das Unwetter	IN	Als an einem Abend der Strom ausfiel, holte Nina aus der Schublade eine Kerze und entflammte ein Streichholz.	Kerze angezündet
Das Unwetter	UN	Als an einem Abend der Strom ausfiel, holte Nina aus der Schublade ein Nachthemd und legte sich schlafen.	Kerze angezündet

H.4. Pseudoword materials

Headline	Pseudoword sequence	Test probe
Pseudowörter	uds wur fobel gumdo qij lev kujs dorbig anbyv wafsuv falnul doj oenim woo mo kohnro seepu jubdo	seepu jubdo
Pseudowörter	cozyr roun cu iyn uiz do freqnong piope lizle ge fiet fulbex ouk ad gug relg vaftel vod	vaftel vod
Pseudowörter	wun xok hatjut gil dukix pow hyk seovo il soanin ounas nuw comnurir cullhe tuz reinuq walqen duf	walgen duf
Pseudowörter	ikt sym deroa luntpemve kasde iz luztu hojüfer uv dirteke koz zew foch goc otbes por wao jujke	ikt sym

H.5. Filler materials Experiment 1

Headline	Pseudoword sequence	Test probe
Der Bäcker	Beim Bäcker war es ungewöhnlich voll. Als Maren endlich an der Reihe war, gab es keine Brötchen mehr.	Brötchen gebacken
Das Festmahl	Das Gemüse war im Topf, das Öl siedete in der Pfanne und das Hähnchen war zum Braten vorbereitet.	Pfanne gespült
Der Safariurlaub	Eine Safari war schon immer Franziskas Traum gewesen. Auf der Fahrt mit dem Jeep machte sie viele Fotos.	Kamera kaputt
Die Sprachkurse	Für Sprachen hatte Doris sich in der Schule nie interessiert. Dafür belegte sie viele Sprachkurse an der Uni.	Sprachkurs ausgefallen

H.6. Filler materials Experiment 2

Headline	Pseudoword sequence	Test probe
Der Bäcker	Beim Bäcker war es ungewöhnlich voll. Als Maren endlich an der Reihe war, wurden gerade neue Brötchen gebacken.	Brötchen gebacken
Die Safari	Eine Safari war schon immer Franziskas Traum gewesen. Auf der Fahrt mit dem Jeep machte sie viele Fotos.	Fotos gemacht
Der Walkman	Viola hatte ihren Walkman auf volle Lautstärke gedreht. Daher konnte sie nicht hören, daß Frank nach ihr rief.	Frank rief
Die Gartenarbeit	Andrea wollte bei der Gartenarbeit nicht ihre neuen, teuren Schuhe ruinieren. Also zog sie lieber ihre Gummistiefel an.	Gummistiefel angezogen